

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 15:58:31 ; Search time 16 Seconds
(without alignments)
1280.550 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068

Sequence: 1 MKKFLIIAAVAFVAVSADPI.....VRKENTKVLAPAFKRELEKN 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR 78:**

1: pir1:**
2: pir2:**
3: pir3:**
4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	10.2	568	2 T49362	hypothetical prote
2	97.5	9.1	776	2 E90591	lipoprotein [impor
3	93.5	8.8	449	2 F84241	hypothetical prote
4	93.5	8.8	628	2 G75154	tungsten-containin
5	92.5	8.7	373	2 T36830	probable phenylala
6	92.5	8.7	697	2 B86239	protein T10024.19
7	91.5	8.6	979	2 E90601	hypothetical prote
8	91.5	8.6	1393	2 E72122	RNA polymerase bet
9	91.5	8.6	1393	2 B86501	RNA polymerase bet
10	91.5	8.6	1397	2 B81548	DNA-directed RNA p
11	90.5	8.5	547	2 F82783	60kDa chaperonin x
12	90.5	8.5	681	2 T15590	hypothetical prote
13	90	8.4	1558	2 B71503	RESA-H3 antigen p
14	88.5	8.3	459	2 A46372	immunophilin FKBP5
15	88.5	8.3	2137	2 T05244	hypothetical prote
16	88	8.2	436	2 B95199	phosphoglycerate d
17	88	8.2	436	2 H98055	phosphoglycerate d
18	87	8.1	372	2 G96818	PgK20.3 [imported]
19	87	8.1	399	2 T32933	hypothetical prote
20	87	8.1	1640	2 D86798	prophage p13 prote
21	86.5	8.1	458	2 JN0873	immunophilin p59 -
22	86.5	8.1	544	2 B82048	chaperonin, 60 kd
23	86	8.1	303	2 D83482	probable transcrip
24	86	8.1	339	2 E70710	hypothetical prote
25	85.5	8.0	903	1 C64444	cell division cont
26	85.5	8.0	1396	2 G71529	DNA-directed RNA p
27	85.5	8.0	1396	2 F81666	DNA-directed RNA p
28	85.5	8.0	1441	2 T00335	hypothetical prote
29	85	8.0	292	2 G75084	hypothetical prote

30	85	8.0	878	2	P86795	cation-transprotin
31	84.5	7.9	1896	1	RNFF2L	DNA-directed RNA p
32	84	7.9	314	2	A72335	conserved hypothet
33	84	7.9	392	2	G89774	phosphopentomutase
34	84	7.9	513	2	G47788	d-xylose transport
35	84	7.9	600	2	T17436	ATP-binding protei
36	84	7.9	600	2	AF0233	inner membrane ABC
37	84	7.9	775	2	G86696	cation-transprotin
38	83.5	7.8	474	2	G97033	beta-glucosidase f
39	83.5	7.8	548	2	AC0043	60 kDa chaperonin
40	83.5	7.8	550	2	S2901	heat shock protein
41	83	7.8	244	2	D64358	ribosomal protein
42	83	7.8	465	2	F97034	aspartyl aminopept
43	83	7.8	633	2	I39585	dnak-type molecula
44	83	7.8	633	2	AE2591	DNAA protein [impo
45	83	7.8	633	2	E97373	dnaj protein (heat
46	83	7.8	1060	2	H86686	DNA-directed DNA p
47	82.5	7.7	279	2	G84435	cobalt transport A
48	82.5	7.7	550	2	S26423	heat shock protein
49	82.5	7.7	566	2	T47934	hypothetical prote
50	82.5	7.7	2339	2	S41121	acetyl-CoA carboxy
51	82	7.7	331	2	T28374	hypothetical prote
52	82	7.7	357	2	H89819	conserved hypothet
53	82	7.7	461	2	G86594	lipoamide dehydrog
54	82	7.7	461	2	C72031	2-oxo acid dehydro
55	82	7.7	513	2	B91185	hypothetical prote
56	82	7.7	513	2	H86031	hypothetical prote
57	82	7.7	636	2	A84252	Htr15 transducer f
58	82	7.7	2104	2	T38774	myosin-3 heavy cha
59	81.5	7.6	547	2	B49203	heat shock protein
60	81.5	7.6	745	1	S75354	probable copper-tr
61	81	7.6	154	2	H82810	bacterioferritin X
62	81	7.6	244	1	R3MX4	ribosomal protein
63	81	7.6	298	2	F69119	cation efflux syst
64	81	7.6	445	2	E70360	glucose inhibited
65	81	7.6	798	2	S29815	N-ras upstream pro
66	81	7.6	1201	2	P86386	hypothetical prote
67	80.5	7.5	311	2	B70451	prophenate dehydro
68	80.5	7.5	374	2	C84040	hypothetical prote
69	80.5	7.5	705	2	T47949	hypothetical prote
70	80.5	7.5	1365	2	T45031	hypothetical prote
71	80	7.5	160	2	T27963	hypothetical prote
72	80	7.5	292	2	E72223	conserved hypothet
73	80	7.5	455	2	F72326	hemolysin-related
74	80	7.5	798	2	S11210	probable unr prote
75	80	7.5	1213	2	T19835	hypothetical prote
76	80	7.5	1487	2	G96827	protein F20Bd7.10
77	79.5	7.4	246	2	F70062	hypothetical prote
78	79.5	7.4	458	1	A42386	hsp 90-binding pro
79	79.5	7.4	488	2	C86183	hypothetical prote
80	79.5	7.4	548	2	AS1045	GroEL protein [imp
81	79.5	7.4	664	2	G98894	protein kinase [im
82	79.5	7.4	698	2	D90541	atp synthase beta
83	79.5	7.4	819	1	C71527	endopeptidase La (
84	79.5	7.4	906	2	A43000	alpha N-catenin -
85	79.5	7.4	1006	2	C70445	ATPase subunit of
86	79	7.4	220	2	A57418	pyruvate kinase -
87	79	7.4	333	2	H82245	glyceraldehyde 3-p
88	79	7.4	338	2	G90203	ethylene-inducible
89	79	7.4	401	2	A33312	ATP synthase beta
90	79	7.4	412	2	AC2152	dolichyl-phosphate
91	79	7.4	486	2	F75045	2-isopropylmalate
92	79	7.4	503	2	C84595	similar to pEARLI
93	78.5	7.4	326	2	T07825	fibritillin homolog
94	78.5	7.4	351	2	I64030	hypothetical prote
95	78.5	7.4	505	2	D89064	isopropylmalate sy
96	78.5	7.4	528	2	S70295	probable membrane
97	78.5	7.4	573	2	AB2401	hypothetical prote
98	78.5	7.4	735	2	T13646	hypothetical prote
99	78.5	7.4	818	2	F89819	endopeptidase [imp
100	78.5	7.4	819	2	E81681	proteinase, Lon fa

ALIGNMENTS

RESULT 1

T49962
Hypothetical protein F8M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49962
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493
A:Accession: T49962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <BEV>
A:Cross-references: EMBL:AL353993; GSPDB:GN00063; ATSP:F8M21.160
A:Experimental source: cultivar Columbia; BAC clone F8M21
C:Genetics:
A:Gene: ATSP:F8M21.160
A:Map position: 5
A:Introns: 200/3; 337/3; 544/3

Query Match 10.2%; Score 109; DB 2; Length 568;
Best Local Similarity 26.7%; Pred. No. 0.73; Mismatches 36; Indels 26; Gaps 8;
Matches 46; Conservative 36; Mismatches 36; Indels 26; Gaps 8;
QY 25 ITEINKAIDDAIAIEQSETIDPMK--VPDHADKFRHVGIVDFKGLAMRN-----IE 77
DB 331 ITISAREVFEDA----YSPTEIAVMRLQPKCSKVERDSGLVSFTTLLVPSSRIGCIL 395
QY 78 ARG---LKMKGQGDANVK--GEEGIVKA-----HLIGVHDDIVSMYDIAYKGLDHLPT 128
DB 386 GKGGAIITEMRMTKANRIILGKLNPKVASDDDMQVNFVLLKFSQFLSLRLYF 445
QY 129 THVISDIOQVVVALS--LEISDEGNITWTSFEVQFANVNVNHHIGLSILDPI 178
DB 446 TWSVNSNSNFDVQISGLVDVAKEALIQITS---RLRNVFDEGAVSALMPV 494

RESULT 2

E90591
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CT1P)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: E90591
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <KUR>
A:Cross-references: GB:AL445566; PID:q14090052; PIDN:CAC13810.1; GSPDB:GN00153
A:Experimental source: strain UAB CT1P
C:Genetics:
A:Gene: MYPV 6370
A:Genetic code: SGC3

Query Match 9.1%; Score 97.5; DB 2; Length 776;
Best Local Similarity 20.1%; Pred. No. 8.3;
Matches 52; Conservative 37; Mismatches 73; Indels 97; Gaps 10;
QY 22 YDKITEINKAIDDAIA---AIEQSETIDPMKVPD-HADKFE-----RHVGIVD--- 66
DB 437 YNQRFDQVKNKLEQAFPKNSWEQTFNPSPDESNTYQDKLDNIGDEDKAKFTGIDELF 496
QY 67 -----FKGELAMRN-----IEARGLKQMKRQGDANVKGEEGIVK-----A 101
DB 497 NELKSLEFKKAIARNAIAATIGLNTLAELKKGESQASRDIVKADFEKEMQHNDNA 556
QY 102 HL-----LIGVHDDIVSMYDIAYKGLDHLPTTHVISDIOQFVVALSLEIS----- 147

Db 557 HLRKIEKYIFGVNGDEQSKDLDLVKQSQDFTSAVESIH---AKALEFSRKYAKNV 612
QY 148 DEGNI-----TWTSFEVQFANVNVH 168
DB 613 DENNIKKAVELYLANVWNOAIWNFKFNKYLEIWKTSFKKKDGQSDSMTEIQVINEVVK 672
QY 169 IGLSILDPFVGLSDVLT 187
DB 673 YRSSNVADKLIISVITSTNT 691

RESULT 3

F84241
Hypothetical protein Vng0849c [imported] - Halobacterium sp. NRC-1.
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84241
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lia
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: F84241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-449 <STO>
A:Cross-references: GB:AE004437; NID:gi0580415; PIDN:AAG19298.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0849C

Query Match 8.8%; Score 93.5; DB 2; Length 449;
Best Local Similarity 21.1%; Pred. No. 8.7;
Matches 42; Conservative 35; Mismatches 87; Indels 35; Gaps 6;
QY 13 VAVSADPHYDKITEINKAIDDAIAIEQSETIDPMKVPDHADKFERHV----- 62
DB 263 VAALADAIRNDVADEDLATIQDAMGGVPFESTRVELTHLQSRVSEIEAYTDALSSFFIDN 322
QY 63 -----GIVDFKGLAMRNIEARGLKQ-MKRGDANVKGEEGIVKAHLIGVHD 109
DB 323 GTAQVDNLDEDDVADLELSTELSTVDDVAGVEQSVDAADRAALHERVDDVEATVGAVD 382
QY 110 DIVSMEYDIAYKGLDHLPTTHVISDIOQFVVALSLEISDEGNITWTSFEVQFANVNVH 169
DB 383 D---LESSILDRLRGVDALDERLADTDDVLAALA-DVS--GDVDALEADVAELSEHVD 436
QY 170 GGLSILDPFVGLSDVLT 188
DB 437 -----DWRDQLSDVFS 449

RESULT 4

G75154
tungsten-containing aldehyde ferredoxin oxidoreductase (aor-1) PAB2085 - Pyrococcus abyssi
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75154
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A:Reference number: A75001
A:Accession: G75154
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-628 <KAM>
A:Cross-references: GB:AJ248284; GB:AL096936; NID:95457730; PIDN:CAB49318.1; PID:9545782
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2085
C:Superfamily: probable aldehyde ferredoxin oxidoreductase aor-4

[illegible]

60kDa chaperonin XP0615 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R;Accession: F82783
R;Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; PMID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-547 <SIM>
A;Cross-references: GB:AE003907; GB:AE003849; NID:g9105484; PIDN:AAF83425.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sancelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XP0615
C;Superfamily: chaperonin groEL

Query Match 8.5%; Score 90.5; DB 2; Length 547;
Best Local Similarity 22.8%; Pred. No. 19;
Matches 54; Conservative 43; Mismatches 87; Indels 53; Gaps 12;
QY 13 VAVSADPIHDKTSEINKAIDA-TAAIEQSEITIDPMKVPDHADKFERHVGIVDFKGL 71
DB 107 VAAQMNP-----DKRGIDKAVIAAATVLEKISK---PTSDDKAIAQVATISANSDE 156
QY 72 AMNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAYKLGDLHPHTTHVISDIQDFVVA 141
DB 157 SIGNIIAEAMKVKGEVITI--EEGTLENEL---DVVEGQDFRGYSPPYFINNQQS 210
QY 121 ---KLGLDHPHTH--VISDIQDFVVALSLEISDEGNITMSTFVRQFA---NVVNHIGGL 172
DB 211 QIVELDNPTLLHDKKISSVRDLTLVLDVAVAKESKPLIIAEEVEGEALATLVVNNIRGI 270
QY 173 ---SILDPTFG---VLSDEL---TAIFOD---TVRKEMTKVLAPAFKRELEK 212
DB 271 IKVCAKAPGFGRRKAMLEDMAVLGTGTVISEVGLSEKATTSHLGKAKKRVSK 327

RESULT 12
T15590
hypothetical protein C24A3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15590
R;Favell, T.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C24A3.
A;Reference number: Z18373
A;Accession: T15590
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-681 <FAV>
A;Cross-references: EMBL:U40424; NID:g1065542; PID:g1065549; PIDN:AAA81461.1; CESP:C24A3
C;Genetics:
A;Gene: CESP:C24A3.1
A;Introns: 35/3; 79/2; 154/3; 193/3; 235/2; 283/3; 321/2; 354/3; 380/2; 422/1; 505/3

Query Match 8.5%; Score 90.5; DB 2; Length 681;
Best Local Similarity 25.1%; Pred. No. 25;
Matches 51; Conservative 41; Mismatches 62; Indels 49; Gaps 11;
QY 24 KITTEINKAIDD---ATAAIE-----QSEIT-----DPMKVPDHADKFERHVGIVDFKGL 71
DB 133 KXTNQKAKDDFNSRISLVLSMIRAQSERIVANENMR-----HSEAKLAELUSGOLEL 187
QY 72 AMNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAYKLGDLHPHTTHV 131
DB 188 AMKOITSEKAKQKER-----FQKVNEAALAEHL--ELGNSKIDKLM 228
QY 132 ISDIQDFVVALSLEISDEGNIT-MTSFEVQFANVNHIGGLS-ILDPFGVLSDV-LTA 188
DB 229 NSEIQ-----ARKLHEKGLAKMTIEDR-----VNNYVGMKMSIDEMKNGKNVEMPA 278
QY 189 IFQDTVRKEMTKVLAPAFKRELE 211
DB 279 LDTALRREMAIAADKKLSME 301

RESULT 13
B71603
RESA-H3 antigen PB0915w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
R;Accession: B71603
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:95021743; PMID:9804551
A;Accession: B71603
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1558 <GAR>
A;Cross-references: GB:AE001424; GB:AE001362; NID:g3845307; PIDN:AAC71972.1; PID:g3845305
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: FFB0915w

Query Match 8.4%; Score 90; DB 2; Length 1558;
Best Local Similarity 23.3%; Pred. No. 75;
Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;
QY 22 YDKITEINKAIDAIAIEQSEITIDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGL 81
DB 747 FNTVLDKVEETVEISGESLENNE---MDKAFFSEIFDNVKGIOENLITGMFRSIEISIV 802
QY 82 QMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAYKLGDLHPHTTHVISDIQDFVVA 141
DB 803 IQSEKVDLN---ENVVSSIL---DNIENKKEGLLNKLENISSTEGVQETVTEHV-- 851
QY 142 LSLEISDEGNITMSTFV-----RQFANVNHIGGLS-----ILDPFGVLSDVLA--IFQ 191
DB 852 -----EQNV--YVDVDVPAMKQDFLGLNLAAGLKEKMFNLEDVFKSESVDITVTEIKD 903
QY 192 DTVRKEMTKVLAPAFKRELEKN 213
DB 904 EPVQKEVEKTVSIIIE-EMEN 924

RESULT 14
A46372
immunophilin FKBP52 - human
N;Alternate names: 56K protein; FK506-binding protein; Hsp56; probable peptidylprolyl cis
C;Species: Homo sapiens (nan)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
R;Accession: A46372; A35841; C42576; A45115
R;Peattie, D.A.; Harding, M.W.; Fleming, M.A.; DeCenzo, M.T.; Lippke, J.A.; Livingston, I
Proc. Natl. Acad. Sci. U.S.A. 89, 10974-10978, 1992
A;Title: Expression and characterization of human FKBP52, an immunophilin that associates
A;Reference number: A46372; MUID:93066366; PMID:1279700

A;Accession: A46372
A;Molecule type: mRNA
A;Residues: 1-459 <PEA>
A;Cross-references: GB:M88279; NID:G186389; PIDN:AAA36111.1; PID:G186390
A;Experimental source: placenta
A;Note: sequence extracted from NCBI backbone (NCBIN:118780, NCBI:118781)
A;Note: part of this sequence was confirmed by protein sequencing
R;Sanchez, E.R.; Faber, L.E.; Henzel, W.J.; Pratt, W.B.
Biochemistry 29, 5145-5152, 1990
A;Title: The 56-59-kilodalton protein identified in untransformed steroid receptor complex.
A;Reference number: A35841; MUID:90335211; PMID:2378870
A;Accession: A35841
A;Molecule type: protein
A;Residues: 'X', 3-14, 'X', 16-21 <SAN>
R;Yem, A.W.; Tomasselli, A.G.; Heinrichson, R.L.; Zurcher-Neely, H.; Ruff, V.A.; Johnson, J. Biol. Chem. 267, 2868-2871, 1992
A;Title: The Hsp56 component of steroid receptor complexes binds to immobilized FK506 and FK506 analogs.
A;Reference number: A42576; MUID:92147620; PMID:1371107
A;Accession: A42576
A;Molecule type: protein
A;Residues: 'X', 3-18 <YEM>
A;Experimental source: Jurkat cells
A;Note: sequence extracted from NCBI backbone (NCBI:80691)
R;Wiederrecht, G.; Hung, S.; Chan, H.K.; Marcy, A.; Martin, M.; Calaycay, J.; Boulton, D. J. Biol. Chem. 267, 21753-21760, 1992
A;Title: Characterization of high molecular weight FK-506 binding activities reveals a novel protein.
A;Reference number: A45115; MUID:93016131; PMID:1383226
A;Accession: A45115
A;Molecule type: protein
A;Residues: 16-32 <WIB>
A;Experimental source: Jurkat cells
A;Note: sequence extracted from NCBI backbone (NCBI:116746)
C;Superfamily: human FK506-binding protein FKBP51; BKBP-type peptidylprolyl isomerase h
F;50-97/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F;319-352/Domain: tetratricopeptide repeat homology <TPR>
F;353-386/Domain: tetratricopeptide repeat homology <TP2>
Query Match 8.3%; Score 88.5; DB 2; Length 459;
Best Local Similarity 24.6%; Pred. No. 22;
Matches 30; Conservative 25; Mismatches 50; Indels 17; Gaps 3;
QY 18 DPVHYDKTEINKAIDDAIAAQSETI-----DPMKVPDHAD-KFERHV 62
DB 75 DKFSFDLKGKGVKAWDIAIATMKVGEVCHITCKPEYAGSAGSPKIPFNATLVPE--V 132
QY 63 GIVDFKGLAMRTEARGKQKQGDANVKGEGIVKAHLIGVHDDIVSMEDVLA YKL 122
DB 133 ELFFPKGEDUTEEDGGIIRIQTRGEGYAKPNEGAVEALGEGYKDLFDQRELAFEI 192
QY 123 GD 124
DB 193 GE 194
RESULT 15
T05244
hypothetical protein F18A5.140 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C;Accession: T05244
R;Beyan, M.; Weber, N.; Grueninger, D.; Schmidheini, T.; Bancroft, I.; Mewes, H.W.; Maye
submitted to the Protein Sequence Database, February 1999
A;Reference number: 215405
A;Accession: T05244
A;Molecule type: DNA
A;Residues: 1-2137 <BEV>
A;Cross-references: EMBL:AL035528
A;Experimental source: cultivar Columbia; BAC clone F18A5
C;Genetics:

A;Map position: 4
A;introns: 61/1; 621/3; 691/3; 1438/3; 1476/2; 1608/3; 1713/3; 1799/3
A;Note: F18A5.140
C;Superfamily: Arabidopsis thaliana hypothetical protein F18A5.140
Query Match 8.3%; Score 88.5; DB 2; Length 2137;
Best Local Similarity 21.8%; Pred. No. 1.4e+02;
Matches 45; Conservative 34; Mismatches 82; Indels 45; Gaps 10;
QY 35 DAIAAIEQSEITDPMKVPDHADK-----FERHVGIVDFKGLAM--RNIEAR-GLKQMKRQ 87
DB 614 DSHFILELVQNDADKNKPEHVEFTLTFILQKTGIVLVNNECGFMPENRALCDVQGSQTKK 673
QY 88 GDANVKGEGI-VKAHLIG-----VHDDIVSMEYDLY-KLGDLPHTT---HVISDIQDF 138
DB 674 GSGGYIGKKGIGKFSVFRVSDAPEIHSNGFHFKFDISEGQIGYILPTVVPHPDIESLSM 733
QY 139 VVALSLEISDEG---NITMTSFEVROFANVNVHGLSILDPIP----- 179
DB 734 LSGRALHLKADGWNCTITLFPRAIDSERTVNIH-----EFMFSDLHPSLLFLHRLQC 787
QY 180 ----GVLSDVLTAFQDVTVRKEMTKV 201
DB 788 IYVRNVLDSDLLVMRKEVVSNNIVKV 813
RESULT 16
B95199
phosphoglycerate dehydrogenase-related protein [imported] - Streptococcus pneumoniae (str
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: B95199
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
n, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I
nson, T.; Hickey, B.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: B95199
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-436 <KUR>
A;Cross-references: GB:AEO05672; PIDN:AAK75787.1; PID:G14973204; GSPDB:GN00164; TIGR:SP46
A;Experimental source: strain TIGR
C;Genetics:
A;Gene: SPI709
C;Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongation
Query Match 8.2%; Score 88; DB 2; Length 436;
Best Local Similarity 21.8%; Pred. No. 22;
Matches 46; Conservative 34; Mismatches 89; Indels 42; Gaps 8;
QY 30 NKAIIDATAAIEQSEITDPMKVPDHADKFERHVGIVDFKGLAMRTEARGKQKQGD 89
DB 21 NRIAGERISIVEDGEGVTRDIYATGEWLNRSFSDMTGG---IDDVAPMEQIKHQAE 77
QY 90 AN-----VKGEGIVK-----AHLIGVHDDIV-----SMEYDL---AYK 121
DB 78 IAMEEADVIVFVSGKEGIDAEVARKLYKTKPVLAVNKNVDNEMRNDIYDFVALG 137
QY 122 LGDLHPTTHV-----ISDIQDFVVALSLEISDEGNITMTSFEVROFANV-----VNHIGGL 172
DB 138 LGEPLFISVHGIGRGDVLDAIVENLPNEYEEFNPDVFKSLIGRPNVGVKSSLLINAILGE 197
QY 173 S---ILDPFGLVSDVLTAFQDVTVRKEMTKV 201
DB 198 DRVIASPVAGTTRDAIDTHFTDQGEFTMI 228
RESULT 17
H98065
phosphoglycerate dehydrogenase [imported] - Streptococcus pneumoniae (strain R6)

Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;
QY 16 SADPHYDKITEEINKAIDDAIAAEQSETIDPMKVPDPHADKFERHVGIVDFKGLAMRN 75
Db 1142 SMSKAQYDEIVKNAQKQDDITSAKKQOT---EVTDKAQK--THDKTVELANSKADKN 1195
QY 76 IEA-----RGLKMKQKQGDANVKGEGIVK-----AHLII----- 105
Db 1196 VKAAAKEQGETVEQYTKFKDNRNLNFIIDGNGVLFHLKMGNGHGVSLKGFATGTR 1255
QY 106 GVHDDIVSMYDLAYKLGDLHPTHTVIS-----DIQDFVVALSL---ETSDEGNITWTS 156
Db 1256 GLAQDETALVGEFGELAH-HPSRGIFAVGQGPPIRNLKAGTSILPHSKKE-FLSLT- 1312
QY 157 FEVRQFANVNHIGLSLTIDPIFGVLSVLTAIFQDTRVKEMTKVL 203
Db 1313 -----ANLPAHADGVS-----GFLSDALGWV--KSTYKDVTSVIS 1345
RESULT 21
JN0873
immunophilin p59 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 16-Jul-1999
C:Accession: JN0873; S38475
R:Schmitt, J.; Pohl, J.; Stunnenberg, H.G.
Gene 132, 267-271, 1993
A:Title: Cloning and expression of a mouse cDNA encoding p59, an immunophilin that associated with the cytoskeleton.
A:Reference number: JN0873; MUID:94040772; PMID:7693550
A:Accession: JN0873
A:Molecule type: mRNA
A:Residues: 1-458 <SCH>
A:Cross-references: ENBL:X70887; NID:9410498; PIDN:CAA50231.1; PID:9410499
A:Experimental source: NIH-3T3 cell
C:Comment: This protein is a part of the cytoplasmic steroid receptor complex and interacts with the cytoskeleton.
C:Superfamily: human FK506-binding protein FKBP51; BKBP-type peptidylprolyl isomerase hF50-97/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F:319-352/Domain: tetratricopeptide repeat homology <TR1>
F:353-386/Domain: tetratricopeptide repeat homology <TR2>
Query Match 8.1%; Score 86.5; DB 2; Length 458;
Best Local Similarity 23.0%; Pred. No. 31;
Matches 28; Conservative 27; Mismatches 50; Indels 17; Gaps 3;
QY 18 DPHYDKITEEINKAIDDAIAAEQSETI-----DPMKVPDPHAD-KFERHV 62
Db 75 DKFSFDLKGGEVKAWDIAVATMKVGEVCHITCKPEYAGNAGSPKIPNATLVFB-V 132
QY 63 GIVDFKGLAMRNIEARGLKMKQKQGDANVKGEGIVKAHLIIGHVHDDIVSMYDLAYKL 122
Db 133 ELFEFKGDLTEBEDGGIIRIRTRGEGYARPNDGAMVEVALEGYHKDLDFDQRELOFEV 192
QY 123 GD 124
Db 193 GE 194
RESULT 22
B82048
chaperonin, 60 kD chain VC2664 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82048
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H. 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: B82035; MUID:20406833; PMID:10952301
A:Accession: B82048
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-544 <HEI>
A:Cross-references: GB:AE004332; GB:AE003852; NID:99657252; PIDN:AAF59805.1; GSPDB:GN0012
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2664
A:Map position: 1
C:Superfamily: chaperonin groEL
Query Match 8.1%; Score 86.5; DB 2; Length 544;
Best Local Similarity 24.4%; Pred. No. 38;
Matches 41; Conservative 21; Mismatches 69; Indels 37; Gaps 5;
QY 6 LIAAFAFAVSA-----DPHYDKITEEINKAIDDAIAAEQSETIDPMKV 51
Db 372 LAGGAVAVIKGAATEVEMKEKKORVEDALHATRAAVEGVAGGVAGVLAIRAAKLUSSIV- 430
QY 52 PDHADKFERHVGIVDFKGLAMRNIEARGLKMKQKQGD-----NVKGEEGIVKAHLII 105
Db 431 ---GDNEEQNVGI-----RVALRAEAPLRQIVKNAGDEESVAVNVRAGEGNYGNAAT 482
QY 106 GVHDDIVSMYDLAYKLGDLHPTHTVISDIQDFVVALSLETSDEGNIT 153
Db 483 GYVGMIEH-----GILDPKTVRSALQFAASVAGLMTTEAVIT 522
RESULT 23
D83482
probable transcription regulator PAL312 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83482
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brladman, S.; Yuan, Y.; Brody, L.L.; Coulter, K.R.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, N.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83482
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <STO>
A:Cross-references: GB:AE004560; GB:AE004091; NID:99947240; PIDN:AAG04701.1; GSPDB:GN0012
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PAL312
C:Superfamily: conserved hypothetical protein H11364
Query Match 8.1%; Score 86; DB 2; Length 303;
Best Local Similarity 22.6%; Pred. No. 21;
Matches 28; Conservative 19; Mismatches 39; Indels 38; Gaps 4;
QY 24 KITEEINKAIDDAIAAEQSETIDPMKVPDPHADKFERHVGIVDFKGLAMRNIEARGLKQ 83
Db 71 KVLDEARLAIDDA-----EHG-----ELRGSURVTTTQBYGLRQ 106
QY 84 MKRQGDANVKGEGIVKAHLIIGV-----HDDIVSMYDLAYKLGDLHPTHTVISDIQ 136
Db 107 L-----VPALQAFARLHPALQVOLSTSLHADLIGERFDFVAILRGLDSTHHAVAL 159
QY 137 DFVW 140
Db 160 SFVW 163
RESULT 24
E70710
hypothetical protein Rv1482c - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70710
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70710
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <COL>
A:Cross-references: GB:279701; GB:AL123456; NID:93261635; PIDN:CAB02032.1; PID:g1524228
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1482c
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1073

Query Match 8.1%; Score 86; DB 2; Length 339;
Best Local Similarity 28.3%; Pred. No. 24;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;

QY 33 IDDAIAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGLKQMKR----- 86
DB 187 LDPAAVADALIQTADLUKADVEPLIERGRGKMAAALDLVGGASPKETWRLLL 246

QY 87 -----QGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKLGLDHPHT--HVISDI 135
DB 247 LIRAGFRPQTQIAVRNMGWAEHLDMGQDIKVAEYD-----GDHLLTSRYHYRKDI 301

RESULT 25
C64444
cell division control protein CDC48 homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: C64444
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: C64444
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-903 <BLU>
A:Cross-references: GB:U67557; GB:L77117; NID:g1591777; PIDN:AAB99153.1; PID:g1591785; T
C:Genetics:
A:Map position: REV1097124-1094413
C:Superfamily: transitional endoplasmic reticulum ATPase; FteH/SEC18/CDC48-type ATP-bind
C:Keywords: ATP; nucleotide binding; P-loop
F:193-400/Domain: FteH/SEC18/CDC48-type ATP-binding domain homology <VAT1>
F:220-227/Region: nucleotide-binding motif A (P-loop)
F:466-674/Domain: FteH/SEC18/CDC48-type ATP-binding domain homology <VAT2>
F:493-500/Region: nucleotide-binding motif A (P-loop)

Query Match 8.0%; Score 85.5; DB 1; Length 903;
Best Local Similarity 25.0%; Pred. No. 85;
Matches 53; Conservative 30; Mismatches 90; Indels 39; Gaps 11;

QY 23 DKITEENKAIID-----ATAAIEQSETIDPMKV-PDHADKFERHVGIVDFKGLAM 73
DB 571 DKVNQLLTDLGNEEPKQDVVIAATNPDIIDPALLRFGRLDRVIL-VVPDEKARLDI 629

QY 74 RNIEARGLKQMKRGDANV-----KGEIGIVKAHL-----LIGVHDDIVSMEYDLAYK 121
DB 630 FKIHTR---SNMLAEDVNLBELAKTKTEGYTGADIEALCREAAMLAVERSI-GKPDVIEVK 685

QY 122 LGDLHPTTHVISDIQDFWALSLEISDEGNIT--MTSFEVRQFANVNVHIGLSILDPIF 179
DB 686 LREL---INVLIQSISGTFRAAAVNLNSVIKATKRESAEAGEFSELKNAIG-----KII 736

QY 180 GVLSDVLTALFQDTPVRKEMTKVLAPAKRELE 211
DB 737 SVLSPAKEKI--EAVEKEIDKFLVINKBELK 766

RESULT 26

F81529

DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain - Chlamydia trachomatis (serotype D, C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 18-Jun-1999
C:Accession: G71529; C36135
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, S.; Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trach
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: G71529
A:Molecule type: DNA
A:Residues: 1-1396 <ARN>
A:Cross-references: GB:AE001304; GB:AE001273; NID:G3328730; PIDN:AAC67907.1; PID:G3328731
A:Experimental source: serotype D, strain UW-3/Cx
R:Engel, J.N.; Pollack, J.; Malik, F.; Ganem, D.
J. Bacteriol. 172, 5732-5741, 1990
A:Title: Cloning and characterization of RNA polymerase core subunits of Chlamydia trach
A:Reference number: A36135; MUID:91008945; PMID:2211507
A:Accession: C36135
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: DNA
A:Residues: 124-128, 'EK', 131-205, 'D', 207-239 <ENG>
C:Genetics:
A:Gene: rpoC
C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain
C:Keywords: nucleotidyltransferase; transcription

Query Match 8.0%; Score 85.5; DB 2; Length 1396;
Best Local Similarity 24.0%; Pred. No. 1.5e+02;
Matches 50; Conservative 33; Mismatches 66; Indels 59; Gaps 12;

QY 43 SETIDPMKVPDHADKFERHVGIVDFKGLAMRNIE-----EARGLKQ-----MKRQ 87
DB 1145 ABLEEARKEPEDAAD-IAKIDGVVDVDFKQKRIILVVRDEITGMEEHLISLTKHLIVQR 1203

QY 88 GDANVKGK---EGIVKAHLII---GVHD--DIVSMEYDLAYKLGLDHPHTTHVISDIQDFV 139
DB 1204 GUSVINGQQLTDGLVVPHEILICVRLEQKVLVNEQVETRLQGV-----DINXK 1255

QY 140 VAL-----SLEISDEGNITM---TSFEVRQFANV---VNHIGG-----LSILD 177
DB 1256 VBIIVRQMLQKVRITDPGDTLLFCEDVDKKEFVEENRRTEDGGKPAQAVPVLGITKA 1315

QY 178 IFGVLSDVLTALFQDTPVRKEMTKVLAPA 205
DB 1316 SLGTESFISAASFQDT-----TRVLTDA 1338

RESULT 27

F81686

DNA-directed RNA polymerase, beta' chain TC0588 [imported] - Chlamydia muridarum (strain C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: F81686
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heigelberg, J.F.; White, O.; Hickey, I. C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1405, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: F81686
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1396 <TET>
A:Cross-references: GB:AS002328; GB:AS002160; NID:g7190627; PIDN:AAF39420.1; PID:g719062
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0588
C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain

Query Match 8.0%; Score 85.5; DB 2; Length 1396;

Best Local Similarity 23.6%; Pred. No. 1.5e+02;
Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12;
QY 43 SETIDPMKVPDHAKFERHVGIVDPKGLAMRNI-----EARGLKQ-----MKRQ 87
DB 1145 AELVEARKPEDAAD-IAKIDGVDFKGIQKNKRILVVRDEVGTGMBEHLISLTGHLIVQR 1203
QY 88 GDANVKG---EGIVKAHLII---GVHD--DIVSMEYDLAYKGLDHPHTTHVISDIDQFV 139
DB 1204 GDSVIKGOQLDGLVVPREIIEICGVRELQKYLNEVQVYRLOGV-----DINDKH 1255
QY 140 VAL-----SLEISDEGNTW---TSPEVROFANV---VNHIGG-----LSILDP 177
DB 1256 IEIIVRQMLQKVRITDPTGDTILLGEDVDKFEFVEENRTEEDGKPAQVPLVLGITKA 1315
QY 178 IFGVLSVLTALFQDTRVRKEMTKVLAPA 205
DB 1316 SLGTESFISAASFQD-----TRVLTD 1338
RESULT 28
T00335
Hypothetical protein KIAA0564 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00335
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00335
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1441 <NAG>
A:Cross-references: EMBL:AB011136; NID:G3043651; PIDN:BA025490.1; PID:G3043652
A:Experimental source: brain; clone HH1811
C:Genetics:
A>Note: KIAA0564

Query Match 8.0%; Score 85.5; DB 2; Length 1441;
Best Local Similarity 22.0%; Pred. No. 1.5e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;
QY 20 IHYDKI--TEENKKAIDAAIEQSETIDPMK-----VDPHAKFERHV-----GIVDF 67
DB 343 LHRDTTGTTLQPSVKDGLIVSDPLVKAIVLGHILVVDKADKAPTNTVTCILKTLVE- 401
QY 68 KGLAM--RNTEARGLKQKQGDANVKGEGIVKAH-----LLIGVHDDIVSMEYDLA 119
DB 402 NEMILADGRIVA-----NSANVNGRENVVVIHPDFRMIVLANRPGFPFLGNDFF 452
QY 120 YKGLDHPHTHVISD-----IQDFVALS--LEISDEGNTWTS 156
DB 453 GTLGDIF-SCHAVDNPKPHSELEMLRQYGNVPPEPILOKLVAAFGELRLADQGIINY-P 510
QY 157 FEVROFANVNHIGGLSILDPFGVLSVLTALFQ-DTRVRKEMTKVL 202
DB 511 YSTREVNIVRHLQKF----PTEG-LSSVVRNVDFDTSYNNDMREIL 552

RESULT 29
G75084
Hypothetical protein PAB0711 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75084
R:Anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: G75084
A>Status: preliminary
A:Molecule type: DNA

A;Residues: 1-232 <KAW>
A;Cross-references: GB:AJ249286; GB:AL096836; NID:G5458366; PIDN:CAB49980.1; PID:G5458492;
A;Experimental source: strain Orsay
C:Genetics:
A;Gene: PAB0711
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1169
Query Match 8.0%; Score 85; DB 2; Length 292;
Best Local Similarity 20.7%; Pred. No. 23;
Matches 44; Conservative 31; Mismatches 62; Indels 76; Gaps 7;
QY 29 INKAIDD-----AIAAIEQSTIDPMKVPDHAKFERHVGIVDFK----- 68
DB 11 INKVIENPKVIIFPLIVLVIFSVPLAFLEKESNIKPLNFEAGVIEKHGAISDMKLPNL 70
QY 69 -----GELAMNIEA-----RGLQKQKQGDANVKGEGIVKAHLIGVHDDIVSNEY 116
DB 71 KLLLLVGLLQLLLSAVQYSIIHYVKTGTGWEAFKGLNENVIOMFL----- 118
QY 117 DLAYKGLDHPHTHVISDIDQFVALSLEISDEGNTWTSFEVROF-ANVNVNHIGL--S 173
DB 119 -----NVISTL-----IVLVAFLIAIFPMPPIVIGGILES 149
QY 174 ILDPFGVLSVLTALFQDTRVRKEMTKVLAPA 206
DB 150 TGAIVFGILLVILSLGIIGSFAIGMTSVIVPAY 182
RESULT 30
F86795
cation-transporting ATPase yoaB [imported] - Lactococcus lactis subsp. lactis (strain ILJ
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: F86795
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A:Reference number: AB6625; MUID:21235186; PMID:11337471
A:Accession: F86795
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-878 <STO>
A:Cross-references: GB:AE005176; PID:GI2724350; PIDN:AAK05464.1; GSPDB:GN00146
A;Experimental source: strain ILI403
C:Genetics:
A;Gene: yoaB
Query Match 8.0%; Score 85; DB 2; Length 878;
Best Local Similarity 24.7%; Pred. No. 90;
Matches 39; Conservative 33; Mismatches 44; Indels 42; Gaps 8;
QY 72 AMNTEARGLKQKQGDANVKGEGIVKA-HLLIGVHDDIVSMEYDLAYKGLDHPHTH 130
DB 106 ALRQMSAPSAKVL-----NGKTSIPARELVVG--DIVSLE-----AGDFIPADG 149
QY 131 VLSIDQDFVALSLEISDEGNTWTSFEVROFANV-----NHIGGSILDPFG 180
DB 150 RLIDVQNLV-----EGMLTGESEPEKESDVEGVALGDRKNVFSLSV--VYG 200
QY 181 VLSVLTALFQDTRVRKEMTKVLAPA-----FKRELEK 212
DB 201 RADFLVTAIEQTEIGKIAQMLETAEAKQTPLOQKLEK 238
RESULT 31
RNPF2L
DNA-directed RNA polymerase (EC 2.7.7.6) II 215K chain [validated] - fruit fly (Drosophila)
C:Species: Drosophila melanogaster
C>Date: 04-Dec-1986 #sequence_revision 31-Mar-1993 #text_change 28-Jul-2000
C:Accession: S04457; A00693; B27677; S60151
R:Jokarst, R.S.; Weeks, J.R.; Zehring, W.A.; Greenleaf, A.L.
Mol. Gen. Genet. 215, 266-275, 1989
A:Title: Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosoph

Query Match 7.8%; Score 83.5; DB 2; Length 474;
Best Local Similarity 23.3%; Pred. No. 55;
Matches 34; Conservative 22; Mismatches 65; Indels 25; Gaps 3

QY 53 DHADKFERHVGIVDFKGLAMRNTEARGLKQMKQEGDANVKGEE--GIYKAHLILGVHD 109
Db 55 DYYHFKEDVALMAENGLKATRFVSWSRYPKGRKVNKLGLEFYKNLVDELLKYNIEP 114
QY 110 DIVSMEYDLAYKGLDLH---PTTHVISDIOQFVVALSLEISDEGNITMTSFVYRQFAN-- 164
Db 115 VLTLYHNDLPQALQDLYGGFSRNVISDFENYCIITLFKNFKDKVYKWTIFNEQNVFTNLG 174
QY 165 -----VVNHIGLS 173
Db 175 YRCAVHPGCIKDKKYLANHIVNLA 200

RESULT 39
AG0043
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AG0043
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prend
delo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Do
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.;
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0043
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-548 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89210.1; PID:gl5978449; GSPDB:GNO0175
C:Genetics:
A:Gene: groEL
C:Superfamily: chaperonin groEL

Query Match 7.8%; Score 83.5; DB 2; Length 548;
Best Local Similarity 18.1%; Pred. No. 66;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIBOSETIDPMKVDPHADKFERHVGIVDFKGL 71
Db 107 VAAGMKNPM-----DLKRGIDKAVIAAVER---LKKLSVPCSDSKAIAQVGTISANSDS 156
QY 72 AMRNTEARGLKQMKQEGDANVKGEGIVKAHLILGVHDD---IVSMEYDLAY----- 120
Db 157 TVGELQAQNEKV-----GKEGVITVEEGSLQDELVDVGGWQFDKGYLSPIFK 207
QY 121 -----KLGDLPHTTHVVISDIOQFVVALSLEISDEGNITMTSFV 159
Db 208 PETGSIELSPFILLADKKISNIREMLPVLEAVAKACKPLLIIAEDVEGEALATL----- 262
QY 160 RQFANVVNHIGL-----SILDPIRGVLSDLVLTAFQTVKEMTKVLAPAFKKELEK 212
Db 263 -----VYNTMRGIVKVAAPKAPGGF---DRKAMQLQDIATLTAGTVISEIGLELEK 311

RESULT 40
S52901
heat shock protein 60K - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
C:Accession: S52901
R:Autenrieth, I.; Noll, A.
submitted to the EMBL Data Library, October 1994
A:Description: Characterization of protective T-cell epitopus of Yersinia ente
A:Reference number: S52901
A:Accession: S52901
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-550 <AUT>

```

RESULT 42
F97034
C:Species: Clostridium acetobutylicum
C:date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97034

```

RESULT 44
AE3591
DNAK Protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE3591
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks,D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

[illegible]

QY	136	-QDFVVALSLEISDEGNITMTSFVRQPVNVNHIGGL-----SILDPFPGVLSDVL	186
Db	452	QBSFIDRLPQPKTKDE-----VIHVGLKDDAMSNTYSLDDVSYGIL--L	492
QY	187	TAIFQDTV	194
Db	493	QIFFSEVV	500
RESULT 50			
S41121			
acetyl-CoA carboxylase (EC 6.4.1.2) - human			
C:Species: Homo sapiens (man)			
C>Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 11-Jan-2002			
C:Accession: S41121			
R:Ra, J.; Daniel, S.; Kong, I.S.; Park, C.K.; Tae, H.J.; Kim, K.H.			
Eur. J. Biochem. 219, 297-306, 1994			
A:Title: Cloning of human acetyl-CoA carboxylase cDNA.			
A:Reference number: S41121; MUID:94139704; PMID:7905825			
A:Accession: S41121			
A:Molecule type: mRNA			
A:Residues: 1-2339 <HAJ>			
A:Cross-references: EMBL:X68968; NID:G452315; PIDN:CAA48770.1; PID:G452316			
C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-			
C:Keywords: biotin binding; ligase			
F:120-620/Domain: biotin carboxylase homology <BCH>			
F:747-819/Domain: lipoyl/biotin-binding homology <LPB>			
F:786/Binding site: biotin (lys) (covalent) #status predicted			
Query Match 7.7%; Score 82.5; DB 2; Length 2339;			
Best Local Similarity 24.2%; Pred. No. 4.7e+02; Indels 41; Gaps 7;			
Matches 39; Conservative 32; Mismatches 49;			
QY	45	TIDPMKVDPHADKFRHVGIVDFKGEL-----AMENIEARGLKQMKRQG-----	88
Db	2091	TINPLCIEMYADK-ESRGGVLEPQGVETVEIKFRKEDLIKSMRRIDPAYKLMQELGEPDLS	2149
QY	89	DANVGEIGVKAHLIGVHDHDIYSMEYDLAYKGLDLHPT-----THVISDIQDFVVAL	142
Db	2150	DKDRKDLGRLKAR-----EDLLPIYHQVAVQAFDHDTEGRMLEKGVISDILEWKTAR	2204
QY	143	S-----LEISDRGNITMTSFVRQPVNVNHIGLSIL	175
Db	2205	TFYWRRLRLLEDQ-----VKQELQASGELSHVHIQSL	2240
RESULT 51			
T28374			
hypothetical protein 213 - Melanoplus sanguinipes entomopoxvirus			
C:Species: Melanoplus sanguinipes entomopoxvirus			
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 16-Feb-2001			
C:Accession: T28374			
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.			
J. Virol. 73, 533-552, 1999			
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.			
A:Reference number: Z20484; MUID:99102612; PMID:9847359			
A:Accession: T28374			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-331 <AFO>			
A:Cross-references: EMBL:AF063866; NID:G4049647; PIDN:AAC97703.1; PID:G4049743			
C:Genetics:			
A>Note: MSV213			
C:Superfamily: Melanoplus sanguinipes entomopoxvirus hypothetical protein 213			
Query Match 7.7%; Score 82; DB 2; Length 331;			
Best Local Similarity 22.5%; Pred. No. 47;			
Matches 45; Conservative 30; Mismatches 53; Indels 72; Gaps 11;			
QY	20	IHYDKITE-EINKAIDDAIA-----AIEQSETIDPMKVPDHADKFRHVGIVDFKGELA	72
Db	62	ICYDKISLSKNSIADKITITPKDNGIYVEELI-----FNNINIIDFK--	106

55 T. L. S. 55

B91185
 hypotetrical protein ECs4450 [imported] - Escherichia coli (strain O157:H7, substrain R1
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 30-Sep-2002
 C/Accession: B91185
 G. Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G.
 Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: B91185
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-513 <HAY>
 A:A/Cross-references: GB:BA000007; PID:BA037873.1; PID:GI3363924; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: ECs4450

Query Match 7.7%; Score 82; DB 2; Length 513;
Best Local Similarity 23.0%; Pred. NO. 80;
Matches 43; Conservative 33; Mismatches 71; Indels 40; Gaps 7;

31 KAIDDAIAIEQSETID-----PMKVDPDHADKFERHVGIVDFKGELAMRNIEAR 79

18 K A I D N V S L R L N A G E T V S L C G E N S G K S T L M K V L C G I Y P H S Y E G E I I F A G E -----E I O A S 73

80 GLKOMKROGDANVKGEIGVK-----AHLIG---VHDDIVSMEYDLAYKLGDLHPPTHV 131

[illegible]

132 ISDIODFVVALSLEISDE---GNITMTSFEVROFANVNVNHIIGLSILDPIFGVLSVLT 188

[illegible]

189 TEODTVR 195

Y
EOT
TFQD
|
:
:

RESIST 56

H56031 36
 H56031
 Hypothetical protein xylG [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H56031
 C:RefName: N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Gruber, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

Accession: H86031
Status: preliminary
Molecule type: DNA
Reference number: H03400, H015:21074355, H10:11200321

Cross-references: GB:AE005174; NID:gl2518305; PIDN:AAG58716.1; GSPDB:GN00145;

Experimental source: strain O157:H7, substrain EDL933
Genetics:

... ..

```
74 *****TATTTTGT AT TTTT TT TTT ENTET CNETITCTI ---MVDL-----M 119
```

132 ISDIODEVVAISLEISDE---GNITMTSFEVROFANVNHIGGI.SILDPIFGVLSDVLT 188
Ov

178	168
179	169
180	170
181	171
182	172
183	173
184	174
185	175
186	176
187	177
188	178
189	179
190	180
191	181
192	182
193	183
194	184
195	185
196	186
197	187
198	188
199	189
200	190
201	191
202	192
203	193
204	194
205	195
206	196
207	197
208	198
209	199
210	200
211	201
212	202
213	203
214	204
215	205
216	206
217	207
218	208
219	209
220	210
221	211
222	212
223	213
224	214
225	215
226	216
227	217
228	218
229	219
230	220
231	221
232	222
233	223
234	224
235	225
236	226
237	227
238	228
239	229
240	230
241	231
242	232
243	233
244	234
245	235
246	236
247	237
248	238
249	239
250	240
251	241
252	242
253	243
254	244
255	245
256	246
257	247
258	248
259	249
260	250
261	251
262	252
263	253
264	254
265	255
266	256
267	257
268	258
269	259
270	260
271	261
272	262
273	263
274	264
275	265
276	266
277	267
278	268
279	269
280	270
281	271
282	272
283	273
284	274
285	275
286	276
287	277
288	278
289	279
290	280
291	281
292	282
293	283
294	284
295	285
296	286
297	287
298	288
299	289
300	290
301	291
302	292
303	293
304	294
305	295
306	296
307	297
308	298
309	299
310	300
311	301
312	302
313	303
314	304
315	305
316	306
317	307
318	308
319	309
320	310
321	311
322	312
323	313
324	314
325	315
326	316
327	317
328	318
329	319
330	320
331	321
332	322
333	323
334	324
335	325
336	326
337	327
338	328
339	329
340	330
341	331
342	332
343	333
344	334
345	335
346	336
347	337
348	338
349	339
350	340
351	341
352	342
353	343
354	344
355	345
356	346
357	347
358	348
359	349
360	350
361	351
362	352
363	353
364	354
365	355
366	356
367	357
368	358
369	359
370	360
371	361
372	362</

189 TENDITZ 195

QY T89 TFQD:VR T95

RESIII.T 57

RESOL1 5/
 A84252
 Htr15 transducer [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:accession: A84252
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Author: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lie
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160: MIMD:205504483: PMID:11016950

Query Match 7.7%; Score 82; DB 2; Length 636;

Query Match 7.7%; Seed 02; Zengun 00
 Best Local Similarity 19.9%; Pred. No. 1e+02;
 Matches 48; Conservative 36; Mismatches 69; Indels

26 TEEINKAIDDAIAAI--EOSETIDPMK-----VPDHADK 57

07	20	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96	94	92	90	88	86	84	82	80	78	76	74	72	70	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38	36	34	32	30	28	26	24	22	20	18	16	14	12	10	08	06	04	02	00	98	96
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

58	FERHVGIVDEK-GELA	-----MRNIE-----	77
----	------------------	-----------------	----

OY		58	FERNANDIN-GEWA	:			PARTIS
DZ		261	BEDI VACITOCOTARI ANDIETATADINTADIIVISCANEMI SCANTEMENEGASMOEVAIS	:			MARIE

78 ----APGLKOMKROGDANVKGEETVKA-HLLTGVHDDTVSMETVDIAYKLGDIHPPTHTVI 132

[illegible]

123 SPTDNEFVAI SLEFSDE-----GNITMTSFEVBOFANVNVHTGGLSLIDPIEGVLS D 184

[illegible]

185 V 185

185
185

BEST.T 58

myosin-3 heavy chain, type II, myp2 [similarity] - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 08-Sep-2000
 C:Accession: T38774; T43276; T43553
 R:Skelton, J.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, August 1997
 A;Reference number: Z21751

A;Accession: T38774

A;Status: preliminary; tr

A;Molecule type: DNA
A;Residues: 1-2104 <SKE>

A;Cross-references: EMBL:Z98762; PIDN:CAB11475.1; A:Experimental source: strain 972h(=): CGmid C4A8

R;Motegi, F.; Nakano, K.; Kitayama, C.; Yamamoto, M.; Mabuchi, I.
FEBS Lett. 420, 161-166, 1997
A;Title: Identification of Myo3, a second type-II myosin heavy chain in the fission yeast
A;Reference number: Z22385; MUID:98119390; PMID:9459302
A;Accession: T43276
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2104 <NOT>
A;Cross-references: EMBL:AB007633; NID:G2828347; PIDN:BAA24579.1; PID:G2828348
R;Bezanilla, M.; Forsburg, S.L.; Pollard, T.D.
Mol. Biol. Cell 8, 2693-2705, 1997
A;Title: Identification of a second myosin-II in *S. pombe*: Myp2p is conditionally required
A;Reference number: Z22572
A;Accession: T43553
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1192, 'G', 1194-1303, 'G', 1305-1343, 'K', 1345-1419, 'D', 1421-2104 <BEZ>
A;Cross-references: EMBL:AF029788; PIDN:AAC04615.1
C;Genetics:
A;Gene: SPAC4A8.05c; myp2
A;Map position: 1
C;Function:
A;Description: may be involved in cytokinesis and stabilization of F-actin cables
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP
F;92-755/Domain: myosin motor domain homology <MMO>

Query Match 7.7%; Score 82; DB 2; Length 2104;
Best Local Similarity 24.8%; Pred. No. 4.5e+02;
Matches 36; Conservative 25; Mismatches 54; Indels 30; Gaps 6;
QY 24 KITEINKAIDDAIAAEIQSETIDPMKVPDHADKPFERHVGIVDFKGEIARLN-----75
DB 1864 EVSEQLNKKYDSSARP-----DEIEMEKEVLTLSNLAQKDLLSSLVE 1908
QY 76 IEARGLQKMKQGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKLGDLHPTTHVSD 134
DB 1909 RIQIEINFALTKQKSNHREELNQLHRLQGV---LQEKKDLLEKLPDLDLKTYPIST 1964
QY 135 IQDFVWALSLEISD-EGNITWTGFE 158
DB 1965 SKD-VRMLQKQISDLEASFAASDIE 1988

RESULT 59
B49203
heat shock protein GroEL - Haemophilus ducreyi
C;Species: Haemophilus ducreyi
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C;Accession: B49203
R;Parsons, L.M.; Waring, A.L.; Shayegeani, M.
Infect. Immun. 60, 4111-4118, 1992
A;Title: Molecular analysis of the Haemophilus ducreyi groE heat shock operon.
A;Reference number: A49203; MUID:93014108; PMID:1356926
A;Accession: B49203
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-547 <PAR>
A;Cross-references: GB:M91030; NID:g148890; PIDN:AAA24961.1; PID:g148892
A;Experimental source: 35000, ATCC 33922
A;Note: sequence extracted from NCBI backbone (NCBIN:114300, NCBIPI:114302)
C;Superfamily: chaperonin groEL

Query Match 7.6%; Score 81.5; DB 2; Length 547;
Best Local Similarity 18.2%; Pred. No. 94;
Matches 43; Conservative 44; Mismatches 82; Indels 67; Gaps 9;
QY 13 VASADPTHYKITEEINKAIDDAIAAEIQSETIDPMKVPDHADKPFERHVGIVDFKGEI 72
DB 107 VAAGNFM-----DLKRGIDKAAV--VEELKATSKCETSKEIQVGTISANSDET 157
QY 73 MRNIEARGLKQMKRGDANVKGEGIVKAHLIGVHD--DIV-SMEYDLAY-----120

cation efflux system protein (zinc/cadmium) - *Methanobacterium thermoautotrophicum* (strain

QY 90 ANVK-GEEGIVKAHL-----LIGVHDDIVSMEYDLAY-----KIGDL 125

```
Db 251 VDPRTGKPPAVVQLRKNKEGTLISLVGFQTKLYTKEQKRVRLPCLNNAVFLGSM 310
QY 126 HPTHV-----ISDIQDFVVALSLI-----SDEGNITM 154
Db 311 HRTFIQSNKVLTHYLNKKKNIFPAGQITGVGYVASSATGILAGINAGRLARGEKL 370
QY 155 TSPEVRGFANVANH-----GGLSLDPFIFGLSDVLTAFQDVTVRKEMTKVLAPAFKREL 210
Db 371 KAPTETMLGALVNVIVTKGELQPMNPFVGLLPLEKKVDRKKRKLMA-----KRAL 424
QY 211 B 211
Db 425 B 425

RESULT 65
S29815
N-ras upstream protein NRU - human
N;Alternate names: N-ras 5'-region protein NRU; protein DKFZp586J0620.1
C;Species: Homo sapiens (man)
C;Date: 02-Dec-1993 #sequence revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: S29815; A60196; T08779
R;Bousadia, O.; Jacquemin-Sablon, H.; Dautry, F.
Biochim. Biophys. Acta 1172, 64-72, 1993
A;Title: Exon skipping in the expression of the gene immediately upstream of N-ras (unr-
A;Reference number: S29815; MUID:93176829; PMID:8439573
A;Accession: S29815
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 67-166 <BOU>
R;Nicolaiew, N.; Triqueneaux, G.; Dautry, F.
Oncogene 6, 721-730, 1991
A;Title: Organization of the human N-ras locus: characterization of a gene located immedi-
A;Reference number: A60196; MUID:91270888; PMID:2052355
A;Accession: A60196
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-103,135-798 <NIC>
R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A;Reference number: Z16471
A;Accession: T08779
A;Molecule type: mRNA
A;Residues: 430-798 <WAM>
A;Cross-references: EMBL:AL050150
A;Experimental source: adult uterus; clone DKFZp586J0620
C;Genetics:
A;Gene: NRU
A;Note: DKFZp586J0620.1
C;Keywords: DNA binding
F;1-798/Product: N-ras upstream protein NRU version II (exon skipping) #status pr
F;1-103,135-798/Product: N-ras upstream protein NRU version I (exon skipping) #status pr

Query Match 7.6%; Score 81; DB 2; Length 798;
Best Local Similarity 22.9%; Pred. No. 1.6e+02;
Matches 36; Conservative 29; Mismatches 56; Indels 36; Gaps 8;

QY 15 VSADPIHYDKITBEINKAIDDAIAAEQSEITIDPMKVPDAD-KFERHVGIV---DFKGE 70
Db 574 VSAEKVNK---THSVNGITBEADPTYSKGKVIPLRSVDPTQTEYQGMIEIIVEEGMKGE 630
QY 71 LAMENIEARGIKOMKRGCDANVKGEGIVKHAHLIGVHDDIVSMVEYDLA----- 119
Db 631 -----VYPFGIVGANKGDCLOKGES--VKFQLCV-LGQNAQTWAINITPLRATVECVK 682
QY 120 -----YKLGDLHPHTHVISDIDFVVALSLISDE 149
Db 683 DQFGFINYEVDGSKLFFHVKEVQD---GIELQAGDE 716

RESULT 66
F86386
hypothetical protein F14G11.4 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: F86386
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86386
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1201 <STO>
A;Cross-references: GB:AE005172; NID:g11560181; PIDN:AAG38123.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 7.6%; Score 81; DB 2; Length 1201;
Best Local Similarity 25.4%; Pred. No. 2.7e+02;
Matches 47; Conservative 18; Mismatches 64; Indels 56; Gaps 9;

QY 19 PIHYDKITEINKAIDDAIAA-----IEQS-ETIDPMKVPDADKFERHVGIV 65
Db 779 PSPFOVIRELDTKAVGDLAAATDLVWMEPCIEGSEVETEDPNPGEADK-----T 831
QY 66 DF-KGELAMRNIEARGIKOMKRGCDANVKGEGIVKHAHLIGVHDDIVSMVEYDLAYKGD 124
Db 832 DIPKNDSDNAAAEEAKEKKSSPKVKKQ-----LVYEIIDD 869
QY 125 LHP-----TTHVISDIQDFVVALSLI-ISDEGNITMTSFEV---RQFANVNVHIGLSIL 175
Db 870 AHPHGFKAKTVLVPD-----VPAQIEVVIRAGNVSYNPLVDVDFSKIELSLNIKGPSVI 925
QY 176 DPFG 180
Db 926 HTLFG 930

RESULT 67
B70451
prephenate dehydrogenase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 03-Mar-2003
C;Accession: B70451
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70451
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-311 <AQF>
A;Cross-references: GB:AE000754; NID:92984047; PIDN:AAC07589.1; PID:g2984048; GB:AE00065;
A;Experimental source: strain VF5
C;Genetics:
A;Gene: tyrA
C;Superfamily: prephenate dehydrogenase/arogenate dehydrogenase, feedback inhibition-inse

Query Match 7.5%; Score 80.5; DB 2; Length 311;
Best Local Similarity 22.9%; Pred. No. 56;
Matches 56; Conservative 34; Mismatches 88; Indels 67; Gaps 11;

QY 2 MKFLIIRAAVAFVAVS-ADPIH-----YDKITEINKAIDDAIAAEQSEITIDPMK 50
Db 30 MQNVLIIVGVGFWGGSFAKSLRRSGFKGIYVDINPESISKAVD--LGIIDEGT----- 82
```

QY 51 VPDHAKFERHVGIVDPKGLAMENIARGIKOMKRO-----GDANVKGEEGIVKAHL 104
Db 83 -----STAKVEDSPDPVMLSSPVRPREAKKLSILSEDAIVT-DQSGVKGKLV 132
QY 105 IGVHDDIVSMEYDLAYKGLDHPHTTHVISDIDQFVVALSLEISDEGN---ITWTSFEVRQ 161
Db 133 -----YDENILGKRFVGGHPAGTKEGSGVEYSLDNLVEGKKVILTPTKTKDK 181
QY 162 FANVNVH-----IGGL-----SILDPFEGVLSD-----VLTAFQDVTVRKMTKVL 202
Db 182 RLKLVKRWEDVGGVVMEYSPMLHDYVFGVVSHLPHAVAFALVDLILHMSTPEVDLKYF 241
QY 203 APAFK 207
Db 242 GGGFK 246
RESULT 68
C84040
Hypothetical protein BH3123 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84040
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84040
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-374 <STO>
A:Cross-references: GB:AP001517; GB:BA000004; NID:gi10175500; PIN:BA06842.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3123
Query Match 7.5%; Score 80.5; DB 2; Length 374;
Best Local Similarity 22.4%; Pred. No. 71;
Matches 49; Conservative 35; Mismatches 74; Indels 61; Gaps 11;
QY 21 HYDKITEI---NKAIDDAIAAIQSEITIDPMKVPDHDKPFERHVGIVDPKGLAMENIE 77
Db 117 HLEKQEKELQKKEKELDEVGKERADKTIQPEK--DQLD-FEKRSANLD---LRQLQIQ 169
QY 78 ARGKQMKR-----QGDAV-----KGEIGIVKAHLIGVHDDIVSM-----EYDLA 119
Db 170 KDGLEEQKDLVWKSRGDGTMTVNDHLQGGQDGMKPVLEIASAGELQGLKISEYD-- 227
QY 120 YKGLDHPHTTHVISDIDQFVVALSL--EISDEGNITWTSF-----EVRQFANV 165
Db 228 -----SLVISEQEVRISSDALPEEWEIGTVKESVYPPSEATLGNDAIGVQIPIT 278
QY 166 VNHIGGLSILDPFQVLSLVDLTAFQDVTVRKMTKVLAP 204
Db 279 VQLSGDQSKLPGYQIMHIVT-----EERTALTVP 309
RESULT 69
T47949
Hypothetical protein F2A19.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47949
R:De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quef
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z24480
A:Accession: T47949
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-705 <DEH>
A:Cross-references: EMBL:AL132962
A:Experimental source: cultivar Columbia; BAC clone F2A19
C:Genetics:

A:Map position: 3
A:Introns: 61/3; 94/3; 135/3; 173/3; 210/1; 231/3; 287/3; 325/3; 372/3; 471/3; 493/3; 561
A:Note: F2A19.170
Query Match 7.5%; Score 80.5; DB 2; Length 705;
Best Local Similarity 24.6%; Pred. No. 1.5e+02;
Matches 41; Conservative 34; Mismatches 65; Indels 27; Gaps 9;
QY 27 BEINKAID-----DAIAAEQSEITIDPMKV-PDHADKPFERHVGIVD-----FK 68
Db 273 KEVNALEKENNELKPKSELEALEERKLTNSKVFPDATESLTHSTLDKEKPESEFP 332
QY 69 G-----ELAMRNTEARGKQKQGDANVKGEEGIVKAHLIGVHDDIVSMEYDLAYKGLD 124
Db 333 GKREMEQSLQRLQ-MDLKETORERD-KARQELKRLKQHLLEKETESEKMDND-SRLIEE 389
QY 125 LHPTT-HVISDIDQFVVALSLEISD-EGNITWTSFEVRQFANVNVH 169
Db 390 LRQTNEYORSQISHLEKSLKQAISSQEDNRLSNDNQIRKLKDTVDL 436
RESULT 70
T45031
Hypothetical protein Y39B68.e [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45031
R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton, J.
raser, A.; Fulton, L.; Gardner, A.; Green, P.; Hawkins, L.; Hillier, L.; Jier, M.; Johnston
B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopra, A.; Saunders, D.
Nature 369, 32-38, 1994
A:Authors: Showkhen, R.; Sims, M.; Smaaldon, N.; Smith, A.; Smith, M.; Sonhammer, E.; St
took, L.; Wilkinson-Sproat, J.; Wohlman, P.
A:Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.
A:Reference number: S43531; MUID:94150718; PMID:7906398
A:Accession: T45031
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <WIL>
A:Cross-references: EMBL:AL132896; NID:G6434440; PIN:CA60910.1; PID:G6434445
A:Experimental source: clone Y39B68
C:Genetics:
A:Map position: 3
A:Introns: 10/1; 37/1; 171/2; 844/2; 1074/2; 1115/3; 1194/2; 1233/3
A:Note: Y39B68.e
Query Match 7.5%; Score 80.5; DB 2; Length 1365;
Best Local Similarity 21.5%; Pred. No. 3.5e+02;
Matches 43; Conservative 33; Mismatches 89; Indels 35; Gaps 8;
QY 27 BEINKAIDDAIAAEQSEITID-PMKVPDHDKPFERHVGIVDPKGL-----AM 73
Db 322 QSIKSLKPIFSAFETLQAVNKLVKVP-----FPHTKFKQKDEFPYIPVYNNSSVDVM 376
QY 74 RNIEARGKQKQGDANVKGEEGIVKAHLIGVHDDIVSMEYDLAYKGLDHPHTTHV 131
Db 377 EKVEGCDHEGRKYSDDYRSKGIEFINNVK-----IVSDHFTAFYFNFTQLDL 426
QY 132 ISDIDQFVVALSLEISDEGNITWTSFEVRQFANV--NHIGLSILDPFQVLSLVDLTAI 189
Db 427 ETDVAAFV--SLNTDINNRTQSEMEVIEIVKNSNNISLTKIQESVKGI-ENQFKAV 483
QY 190 FQDVTVRKMTKVLAPAFKRE 209
Db 484 RLDSLKNDVFSKISDVFRKD 503
RESULT 71
T27963
Hypothetical protein ZK673.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Feb-2000
C:Accession: T27963

```

C/Accession: F72326
R/Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Dodson, M.L.; Haft, D.H.; Hickey,
Garrett, M.E.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.Nature
Name 399, 323-329, 1999
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A/Reference number: A72200; MUID:9287316; PMID:10360571
A/Accession: F72326
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-455 <ARN>
A/Cross-references: GS:AE001751; GS:AE000512; NID:G4981371; PIDN:AAD35927.1; PID:G4981371
A/Experimental source: strain MSB8
C/Genetics:
A/Gene: TM0845
C/Superfamily: hypothetical protein H10107

Query Match          7.5%; Score 80; DB 2; Length 455;
Best Local Similarity 19.1%; Pred. No. 98;
Matches 43; Conservative 49; Mismatches 79; Indels 54; Gaps 10;

QY 25 ITEINKAIDDAIAAEQSEFTIDM-----KVPDADKFERHVGIVDFKGLAM- 73
DB 213 VKETMPRVD--IVAIEENQTKOLIEVDEGYSRIPVYKETIDNIVGICYAKDVLMSL 270
QY 74 -----RNIEARGLQMKRQG-----DANVKBEGIVKA---HLLIGVHD-----DIVSME 115
DB 271 AKDCEEVKSMKVKDIMREALYVPEWNIIDELTKLKARKIHIAIVDVEGVTAGIVTLE 330
QY 116 -----YDLAYKGLDLPHTHVI---SDIQFVVALSLEISDEGNITWTSF 157
DB 331 DIIEELFGNIMDEYDYDEISGIRKIDERTYIVDGGATPINDIEMELRVQFPETEYETIAGY 390
QY 158 EVROPANVNHIGLSLDLPFGVLSDVLTAFQDVTVRKEMTKVL 202
DB 391 LLEHFKRIPN-VGE-----EAVIGNLYFKVLAVGKNRIEKMVKIL 430

RESULT 74
S11210
probable unr protein - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C/Accession: S11210
R/Jeffers, M.; Paciucci, R.; Pellicer, A.
Nucleic Acids Res. 18, 4891-4899, 1990
A/Title: Characterization of unr; a gene closely linked to N-ras.
A/Reference number: S11210; MUID:90370473; PMID:2204029
A/Accession: S11210
A/Molecule type: mRNA
A/Residues: 1-798 <JBF>
A/Cross-references: EMBL:X52311; NID:G57454; PIDN:CRA36549.1; PID:G57455
C/Keywords: DNA binding

Query Match          7.5%; Score 80; DB 2; Length 798;
Best Local Similarity 22.9%; Pred. No. 28+02; Indels 36; Gaps 8;
Matches 36; Conservative 29; Mismatches 56; Indels 36; Gaps 8;

QY 15 VSADPIHYDKITERINKAIDDAIAAEQSEFTIDPMKVPDHAD-KFERHVGIV---DFKGE 70
DB 574 VSAEKVNK---THSVNGITEEANTPIYSGKVIPLRGVDPTQIEYQGMIEIVEEGDMKGE 630
QY 71 LAMNIEARGLKQMKRQGDANVKBEGIVKAHLIGVHDDIVSMEYDLA----- 119
DB 631 -----VYFPGFIVGANKGDCILQKGES--VKFQLCV-LGQNAQTMAYNITPLRRATVECVK 682
QY 120 -----YKLGDLHPHTHVIDSIQDFVVALSLEISDE 149
DB 683 DQGFQINYEVDGSKLFFHVKEVD--GIELQAGDE 716

RESULT 75
T19835

```

A;Cross-references: GB:A80051173; NID:g7715604; PIDN:AAF68122.1.; GSPDB:GN00141
C;Genetics:
A;Gene: F20B17.10
A;Map position: 1

Query Match 7.5%; Score 80; DB 2; Length 1487;
Best Local Similarity 22.9%; Pred. No. 4.2e+02;
Matches 52; Conservative 35; Mismatches 72; Indels 68; Gaps 13;

Qy 33 IDDAIAAEQSEITDPMKVPDHADKFERHVGIVDFKGELAMRNIEARGLKOMKRG----- 88
Db :
453 VDGRIVAVKSKVVD-----EDKLEEFINEVVILSQINHRNI-----VKLIGCCLE 498

Qy 89 -DANVGEEGIVAKHLIYGDDIVSMEX-----DLAYKLGDH-----PTT 129
Db :
499 TDVPILVYFIPNGNLFEHLHDD--SDDYTTTWVEVRLRAVDIAGALSYLHSAASSPIY 556

Qy 130 HVIASIQDFVALS-----LEISDEGNTIMTSFEVRQPANVNHIGLSILDPP----- 177
Db :
557 H-RDIKSTWMLDKHRAKVDSFGTSRTVTVDHTHLTTVVS--GTGVYMDPEYFQSQQP 612

Qy 178 -----IFG-VLSDVLTAIFQTV-----RKMTKVLAPAFKRELEKN 213
Db :
613 TDKSDVYSFGVLAELITG--EKSVSLRSCQRYRTLATYFTLAMKEN 657

RESULT 77
F70062
hypothetical protein ywmB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: F70062
R;Kunost, P.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte-
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.;
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen-
ier, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadai, Y.; Sato, T.; Scanlon
Kieger, M.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serio
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I.
A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A9580; MUID: 98044033; PMID: 9384377
A;Accession: F70062
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-246 <KUN>
A;Cross-references: GB:F299122; GB:AL009126; NID:g2636029; PIDN:CAB15694.1; PID:g2636022
A;Experimental source: strain 168
C;Genetics:
A;Gene: ywmB
C;Superfamily: Bacillus subtilis hypothetical protein ywmB

Query Match 7.4%; Score 79.5; DB 2; Length 246;
Best Local Similarity 21.8%; Pred. No. 51;
Matches 47; Conservative 38; Mismatches 94; Indels 37; Gaps 9;

Qy 5 LLTAAVAFVASADPHYDKITEINKAIDATAAEQSEITDPMKYPDHADKFERHVG 64
Db :
11 IISVMLSFVIAVFETHASELT-----PLAQMAEGRQDV--SIDKWTLHA-- 55

Qy 65 VDFKEGLAMRNIE----ARGLKOMKROGDANVKEGIVAKHLIIGHVDDJ---IVSMYED 117
Db :
56 ---KQNLSLTEKFYQVKQLKQRYQYDWVIAREDKWKA---IGTYTKNRTSFRLQ 109

Qy 118 LAVKLGDLPHTTHVIDIQFVVALSLEISDEGNTIMTSFEVRQPANVNHIGLSILD 177
Db :
110 LVTTLLKHNPSTLYLXE-----QMSLETSDSNWDYEQFE--RETGLFOE--KWLIFT 160

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of metacillin-resistant *Staphylococcus aureus*.
A'Reference number: A99758; MUID:21311952; PMID:11418146
A'Accession: G89894
A>Status: preliminary
A'Molecule type: DNA
A'Residuals: 1-664 <KUR>
A'CROSS-references: GB:BA000018; PID:g13701020; PIDN:BA42315.1; GSPDB:GN00149
A'Experimental source: strain N315
C'Genetics:
A'Gene: SA1063

Query Match 7.4%; Score 79.5; DB 2; Length 664;
Best Local Similarity 21.6%; Pred. No. 1.7e+02;
Matches 35; Conservative 31; Mismatches 51; Indels 45; Gaps 5;

QY 1 MMKELLIAAFAVAVSADPIHYDKITEINKAIDDAIAEQSETIDPMKV----- 51
DB 355 IFSLLMIALVSVAMVPGNKYETPDVIGSVKEA-----EQIPKNNLGLKISRYSYD 410
QY 52 -----PDRADKFER--HVGIVDFKG-----ELAMRNIEARGLQK 84
DB 411 KYPENELIKTPTGTERVERGSDVDVLSKGPVKVMENVIGLPKEALQKLSGLKDV 470
QY 85 KROGDANVKGEGI-----VKAHLIGVHDDIVSMYVDLAYK 121
DB 471 TIEKYNNQAPKGYANQSVTANTEIAIHDSNKLKYESLGK 512

RESULT 82

acp synthase beta chain [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C'Species: Mycoplasma pulmonis
C'Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C'Accession: D90541
R'Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A>Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*.
A'Reference number: A99512; MUID:21267165; PMID:11353084
A'Accession: D90541
A>Status: preliminary
A'Molecule type: DNA
A'Residuals: 1-698 <KUR>
A'CROSS-references: GB:AL445566; PID:g14089649; PIDN:CAC13409.1; GSPDB:GN00153
A'Experimental source: strain UAB CTIP
C'Genetics:
A'Gene: MYPV 2360
A'Genetic code: SGC3

Query Match 7.4%; Score 79.5; DB 2; Length 698;
Best Local Similarity 23.0%; Pred. No. 1.8e+02;
Matches 46; Conservative 31; Mismatches 80; Indels 43; Gaps 7;

QY 24 KITEINKAIDDAIAEQSETIDPMKVDPHADKFERHVGIVDPKGLANR---NIEARG 80
DB 155 KIFPENKTKQSLIAAIAEQ-----YTDDFR-----KSEINKPRAKFKARN 198
QY 81 -LKQMKROGDANVKGEGIVKAHLIGVHDDIVSMYVDLAYKGLDHPHTHVISDIO-DF 138
DB 199 PLKPLED-----KAYVLEGIVVEKSVQVKIKLTVKEPIINAIPEIQTEQ 245
QY 139 VVALSLEISDEGNITWTS-----FEVRQFANVNVHIGGSLIDPFGVLSVLTAI 189
DB 246 QOTRLLEISLSDSLVAGYVGLREGQIEIGSFARSKNNPYSIPSEKLLGIIDPVGRI 305
QY 190 FQDITVRKMTKVLAPAKRE 209
DB 306 LDDPTHPLVGQYAPMIEE 325

RESULT 83

C71527
endopeptidase La (EC 3.4.21.53) - *Chlamydia trachomatis* (serotype D, strain UM3/Cx)
N'Alternate names: ATP-dependent proteinase Lon; ATP-dependent serine proteinase La
N'Contains: adenosinetriphosphatase (EC 3.6.1.3)
C'Species: *Chlamydia trachomatis*
C'Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Jun-2003
C'Accession: C71527
R'Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, S.; Science 282, 754-759, 1998
A>Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*.
A'Reference number: A71570; MUID:9900809; PMID:9784136
A'Accession: C71527
A>Status: preliminary
A'Molecule type: DNA
A'Residuals: 1-819 <ARN>
A'CROSS-references: GB:AE001307; GB:AE001273; NID:g3328757; PIDN:AAC67939.1; PID:g3328764
A'Experimental source: serotype D, strain UM-3/Cx
C'Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP
C'Genetics:
A'Gene: lon
C'Superfamily: ATP-dependent Lon protease
C'Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine
F;392-399/Region: nucleotide-binding motif A (P-loop)
F;455-460/Region: nucleotide-binding motif B
F;724/Active site: Ser #status predicted

Query Match 7.4%; Score 79.5; DB 1; Length 819;
Best Local Similarity 21.1%; Pred. No. 2.2e+02;
Matches 51; Conservative 35; Mismatches 73; Indels 83; Gaps 10;

QY 16 SADPIHYDKITEINKAIDDAIAEQSETIDPMKVDPH----- 54
DB 6 NTDSONLDPNASEVEKLDSEAEAE-----KVDDHTPPSELPILNKRPFPGMA 57
QY 55 -----ADKFERHVGIVDFKGLAMRNIEARGLQKMKROGDANVKGEG 97
DB 58 APLLIEAGPHYEVLLTLAKSKQKHGLVLTCKEDA--NTLVKGFNQLHRVG----- 106
QY 98 IVKAHLIGVHDDIVSMYVDLAYKGLDHPHTHVISDIOQDFVALSLEISDEGNITWTSF 157
DB 107 -VSAIL-----RINPIEGGSAQVLLSIEDRIRIVKPIQDKYLKAKSVYHKE-NKELTE- 158
QY 158 EVRQFA-NVNVHIGGSLIDPFI-----GVLSVLTIFAFTVRKEMTK 200
DB 159 ELKAYSISIVSIKDLKLNPLFKEELQIFLGHSDFTPEFKLADFSVALTTAT-REELQE 217
QY 201 VL 202
DB 218 IL 219

RESULT 84

A43000
alpha N-catenin - chicken
C'Species: Gallus gallus (chicken)
C'Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C'Accession: A43000
R'Hirano, S.; Kimoto, N.; Shimoyama, Y.; Hirohashi, S.; Takeichi, M.
Cell 70, 293-301, 1992
A>Title: Identification of a neural alpha-catenin as a key regulator of cadherin function.
A'Reference number: A43000; MUID:92346716; PMID:1638632
A'Accession: A43000
A>Status: preliminary
A'Molecule type: mRNA
A'Residuals: 1-906 <HIR>
A'CROSS-references: GB:D111090; NID:g222787; PIDN:BAA01863.1; PID:g222788
A'Experimental source: embryonic brain
A'Note: sequence extracted from NCBI backbone (NCBI:109985, NCBI:109986)
C'Superfamily: alpha-catenin; vinculin amino-terminal homology; vinculin carboxyl-terminal
C'Keywords: cytoskeleton
F;19-267/Domain: vinculin amino-terminal homology <VINN>
F;375-859/Domain: vinculin carboxyl-terminal homology <VINC>

Db 135 NDHIFNPEQHRIVTAASCTTNCIAPVVKVIEHKLGAQASFTTHLNTTOTILDAPHKD 194
QY 142 -----LSLEISDEGNITWTSFEVQFANVNVHIGLSILOPIFGV-LSVDVLTALFQD 192
Db 195 LRRARACGMSUPTTTSKAKI-----IEIFPLKQKIDGHAVRVPLANASLTDIIFDVQD 251
QY 193 TVRKEMTKVLAPAFKRELE 211
Db 252 TTVEEINCLLKQASENELK 270

RESULT 88
G90203
ethylene-inducible protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: G90203
R:She, Q.; Singh, R.K.; Confalonieri, P.; Zivanovic, Y.; Allard, G.; Aweyez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90203
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <KUR>
A:Cross-references: GB:AB006641; NID:gl13813736; PIDN:AAK40886.1; GSPDB:GN00155
C:Genetics:
A:Gene: SS00570
C:Superfamily: hypothetical protein H11647

Query Match 7.4%; Score 79; DB 2; Length 338;
Best Local Similarity 24.8%; Pred. No. 82;
Matches 37; Conservative 22; Mismatches 46; Indels 44; Gaps 7;

QY 13 VAVSADPHYDKITEINKAIDDAIAA-----IEQSEIDPMKVP 52
Db 100 VARWADP-----KITEVNSSITIPVMAKVIGHYVEAKLEALGVMDISEVLTADDE 155

QY 53 DHADKFERHVGIVDFK---GELAVRNIEARGLKQKQGDANVKGEGIVKAHLIGVHD 109
Db 156 HHINKWFSPVFNAGRNLEALRTSE--GASMRKTEAGTGNVSEAVK-HMKI-INS 211

QY 110 DIVSM-----EYDLAYKLGDL 125
Db 212 EIRLSISMSEEDRVKAREYQVPYQIVEL 240

RESULT 89
AI3312
ATP synthase beta chain/transcription termination factor rho BMEI0487 [imported] - Brucel
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AI3312
R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AI3312
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51568.1; PID:gl17982399; GSPDB:GN00190
C:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0487
A:Map position: I

Query Match 7.4%; Score 79; DB 2; Length 401;
Best Local Similarity 25.5%; Pred. No. 1e+02;
Matches 59; Conservative 20; Mismatches 82; Indels 70; Gaps 13;

QY 12 FVAVSADPHYDKI--TEEINKAI-----DDAAIAIEQSE--TIDP 48
Db 180 FADIPADIVHGLILVLTNELDAIPFRQFVADGRFVERMIALNEQDFQFVSGAGGIDP 239
QY 49 MKVP-DHADK-----FERHVGIVDFKGLAMRNIEARGLKQKQGDANVKGEGIVKAH 102
Db 240 ALLPKDHVKAEGAFEAAPARTALMQBIASRIATRG-----AALNIDYGH 286
QY 103 LLIGVHDDIVSM---EYDLAYKLGDLHP-----TTHVISDI-QDFVVALSLEISDEGNIT 153
Db 287 LSSGFGDTLQAMLKQAYDDVF-----AHGVSADLTHSHVDILQKTAKACGCKTG-----T 337

QY 154 MTSFEVQFANVNVHIGLSILDPITFGVLSVLTALTAIFQDVTVRKEMTKVLAP 204
Db 338 MTQGEFLAMGLVDRAGL-----GAGKD--AAFGKIQDVERLAAP 378

RESULT 90
AC2152
dolichyl-phosphate-mannose synthase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2152
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2152
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074469.1; PID:gl17131863; GSPDB:GN00179
C:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2770
C:Superfamily: stress response protein csbB

Query Match 7.4%; Score 79; DB 2; Length 412;
Best Local Similarity 21.0%; Pred. No. 1e+02;
Matches 41; Conservative 36; Mismatches 74; Indels 44; Gaps 8;

QY 15 VSADPHYD-----KITEINKAIDDAIAIEQSEIDPMKVPDHADKFERH 61
Db 29 VTGQPIQFSLIPIPTVKESGNTNRNVIERSQLDDEFGDYELIVVDD-DSPD--GTWEVA 85

QY 62 VGI-VDFKGLAMRNIEARGLKQKQGDANVKGEGIVKAHLIGVHDHDDIVSMEYDLAY 120
Db 86 LSLTAEPQLRVMRQEEGLSTAVIRGWVARG-----ILGVID----- 126

QY 121 KLGLDHPHTHVIDIQDFV-----VALSLEISDEGNITWTSFEVQFANVNVHIGLSIL 175
Db 127 --GDLQHPHYLLLELLSKHHRGADLAVASRHDVGGVGSWSF-IRRFSLRGAQLLGLVIL 183

QY 176 DPFGVLSVLTALTAIF 190
Db 184 PSVLGRVSDPMMSGYF 198

RESULT 91
A75045
2-isopropylmalate synthase (leua-1) PAB0890 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: A75045
R:Anonymous, Genoscape
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A:Reference number: A75001
A:Accession: A75045
A>Status: preliminary

A;Residues: 1-406 <RAW>
A;Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50254.1; PID:g54587
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: leuA-1; PAB0890
C;Superfamily: 2-isopropylmalate synthase leuA

Query Match 7.4%; Score 79; DB 2; Length 486;
Best Local Similarity 21.0%; Pred. No. 1.3e+02;
Matches 58; Conservative 49; Mismatches 64; Indels 106; Gaps 18;

QY 12 FVAVSADPHYD---KITEINKAIDAIAAEQESETIDPMKVDPHADKFERRHVGIVDF 67
DB 98 FIATS--PIHKVKLRKEPEVKKL--AVKAVE-----HATKYTED----VEF 137

QY 68 KGLARN-----IEARG-----LQMKEQGNDANKGEE 96
DB 138 SAEDATESDNWFLVEVVEVAVIDAGATVINPDVTGVATPEEFVELRYLRNISNIKG-- 195

QY 97 GIWKAHLLIGVDDI-VSMYEYL-AVKLG-DLHPHTTHVISD-----IQDFVALSL-- 144
DB 196 ----VOISVECHDDLGLAVNSLSAIRAGAQDEVTNVNGIGERAGNAALEEVIADVRR 251

QY 145 ---EISDEGNITWTSPFROFANVNHHIGLSILDPIFGVLSDVLT-----IFDTVRKE 197
DB 252 DFYKVKTDVNLK---BIARTSKLVSHLTGIET-PNNKAIVGGNAFAHESGIHQDGVLKE 306

QY 198 MT--KVLAP-----AFKRELEK 212
DB 307 RTTYEIIPKKLPFGSGSKI VLGHKSGRHAFRKLLEE 342

RESULT 92
C84595
similar to PEARLI 4 protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84595
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shen, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
J.; et al., 1999
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84595
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-503 <STO>
A;Cross-references: GB:AEO02093; NID:g4803948; PIDN:AAD29821.1; GSPDB:GN00139
A;Gene: Atg20950
A;Map position: 2

Query Match 7.4%; Score 79; DB 2; Length 503;
Best Local Similarity 22.4%; Pred. No. 1.3e+02;
Matches 50; Conservative 40; Mismatches 77; Indels 56; Gaps 12;

QY 13 VAVSADPHYD-KITEINKAIDAIAAEQESETIDPMKVDPHADKFERRHVGIVDFKEL 71
DB 256 VGSSTPWHYKQT-----SVTEYEQRFEPPELDSEQSFTMQ-----EI 298

QY 72 A-MENIEARGLMKRKQ-----GDANYKBEGIVKA-HLLIGVHDDIVSMEXDLAYKL 123
DB 299 ARMRLQSGYKEETQLASETVVSVANVYKVMISAATLEAIIDKHGLAA-----SSKL- 352

QY 124 DLHPHTTHVISDIODPVVALSI EISEGNITWTSPFROFANVNHHIGLSILDFPGVLS 183
DB 353 ---OSTSTRSFYLESLAAAVNMELKSTALRDLTKTFRVAIEAAVAKMDMSVKI-----DVS 403

QY 184 DVLTAIFO-----DT---VRKEMTKVLAFAFKRELEK 212
DB 404 WLKTAVTLAEAVEYGYKDTAKIVRECDREMT-AGKENEE 445

Qy 114 MEYDLAYKLGDLHPTTHV--ISDIQDFVALSL---EISDEGNITWT 155
:
Dd :
VT-----NYAPSEMIKOISDVNIPVAISLRTGEVGEGKLNP 149
:
:
:

RESULT 95

D69064
isopropylmalate synthase - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
C:Accession: D69064
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Alredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.; S., Church, G.M.; Daniels, C.D.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function A;Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: D69064
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-505 <MT>
A:Cross-references: GB:A5000909; GB:A5000666; NID:g2622596; PIDN:AAB85956.1; PID:g2622595
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1481
C:Superfamily: 2-isopropylmalate synthase leua

Query Match **7.4%**; Score 78.5; DB 2; Length 505;
Best Local Similarity **20.6%**; Pred. No. 1.5e+02;
Matches 40; Conservative 36; Mismatches 73; Indels 45; Gaps 8;

Qy 30 NKAI--DDATA-----AIEOSEIIDPM--KVDPHADK--FERHVGIWDFKGELAMR 74
||| |
::: ||| |
::: ||| |
::: ||| |
::: ||| |

Dd 284 NXAIVGNAFAHEAGIIHHGVLEKAETYEPIEMGWCHERIVGLKHGTANALRSKLQEY 343
||| |
::: ||| |
::: ||| |
::: ||| |
::: ||| |

Qy 75 NIEARG-----LKMKRKQGDNV-----KBEGIVKAHLIGVDHDIV 112
||| |
::: ||| |
::: ||| |
::: ||| |
::: ||| |

Dd 344 GIENKEBQFCTLYEQVKRLDGKKRIATDADLRAMAVTILGKSAREIVKLEGIAVMTGESV 403
||| |
::: ||| |
::: ||| |
::: ||| |
::: ||| |

Qy 113 SMEYDLAYKLDLHPPTHV--ISDIQDFVALSLSIDEQNITMTPSVQRFPANVVNHIG 170
||| |
::: ||| |
::: ||| |
::: ||| |
::: ||| |

Dd 404 MPRTAVIKIRIGDEVKTSMWGVPDAINAOISLVSETADIDELNEYEAIT-----G 457
||| |
::: ||| |
::: ||| |
::: ||| |
::: ||| |

Qy 171 GLSIDLPFGVLSD 184
||| |
::: ||| |
::: ||| |
::: ||| |

Dd 458 GTNALAEVFVMSD 471
||| |
::: ||| |
::: ||| |
::: ||| |

RESULT 96

S70295
probable membrane protein YAL028w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 19-Apr-2002
C:Accession: S70295; S51993; S51992
R:Vo, D.

Submitted to the EMBL Data Library, April 1996
A:Reference number: S70291
A:Accession: S70295
A:Molecule type: DNA
A:Residues: 1-528 <VO>
A:Cross-references: EMBL:U12980; NID:g1326053; PID:gl326061; MIPS:YAL028w
R:Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel
submitted to the EMBL Data Library, August 1994
A>Description: The sequence of chromosome I of Saccharomyces cerevisiae.
A:Reference number: S51956
A:Accession: S51993
A:Molecule type: DNA
A:Residues: 167-528 <BUS>
A:Cross-references: EMBL:U12980; MIPS:YAL028w
A>Note: This sequence has been revised in reference S70291
C:Genetics:
A:Cross-references: SGD:S0000026
A:Map position: 1L

C;Keywords: transmembrane protein
F;505-521/Domain: transmembrane #status predicted <TM1>

Query Match 7.4%; Score 78.5; DB 2; Length 528;
Best Local Similarity 23.0%; Pred. No. 1.5e+02;
Matches 41; Conservative 27; Mismatches 65; Indels 45; Gaps 8;

Qy 39 AISOSETIDPMKVP-----DHADKFERHVGIVDFKGLAVRNIEARGLKQM 84
Db SFSQSVAVDPLEPPGNIYSSSNLSNLSDELVDYQRHIG-----LQLQQTALLKHSL 380
Qy 85 KRQGDANVKEEGIVKAHLLIGVHDDIVSMEYDLAYKGLDLHPTHTHVISDI--QDFVVAL 142
Db 381 K--DEVLKDNDLVKN---IANFDKIVK-----ELRDLRSRTIGKWEELVEEDLYMNL 427
Qy 143 SLSEISDEGNITMTSFVROFANVNHIGLSILDPIFGVLSDLVLTAFQDVTVRKEMTK 200
Db 428 KQDFDKE--NPESFEARLSDTINTVAKLQDLEKEMASCKDRLAS-----RKEVMR 476

RESULT 97
AB2401
hypothetical protein all4762 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2401
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.; DNA Res. 8; 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2401
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-573 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA076461.1; PID:gi7133899; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all4762

Query Match 7.4%; Score 78.5; DB 2; Length 573;
Best Local Similarity 24.1%; Pred. No. 1.7e+02;
Matches 51; Conservative 36; Mismatches 84; Indels 41; Gaps 10;

Qy 23 DKITEINKAIDDAIAAIBOSETID---PMKVDPHADKF--ERHVGIVDFKGLAMRN- 75
Db 358 DKALAEID-AIEARIEALNKLRQIENQILEMQQEDSLNKIAHEVHKNIKTRSLLAQRND 416
Qy 76 IEARGLKQMKRGDANVKEEGIVKAHLLIGVHDDIVSMEYDLAYKGLDLHPTHTHVISDI 135
Db 417 AVAQRCMQEVMKGLANLQKRLDKYFL---HNEAFSL-LELADSVQRFLLIQADAVSPI 471
Qy 136 -----QDFVVALSLEISDEGNITMTSFVROFANVNHIGLSILD----- 176
Db 472 PSNQPEWKIRCDQDFILLIS-GVSDDIRIGATFWLRPFSLTLV-GLSLVIGISGLYVDN 529
Qy 177 -----PIFGVLSDLVLTAFQDVTVRKEMTKV 201
Db 530 GLTFGARPPSDYGLILGWSADVASRNLSNL 561

RESULT 98
T13646
hypothetical protein 95B7.9 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13646
R;Ferraz, C.; Vidal, S.; Brun, C.; Bucheton, A.; Demaille, J.G.
Submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17694
A;Accession: T13646

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-735 <PER>
A:Cross-references: EMBL:AL021728; PIDN:CAAL6821.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0000376
A:Introns: 39/3; 93/2; 294/2; 387/2
A:Note: EG:95B7.9
C:Superfamily: fruit fly hypothetical protein EG_95B7.9

Query Match 7.4%; Score 78.5; DB 2; Length 735;
Best Local Similarity 24.7%; Pred. No. 2.3e+02;
Matches 41; Conservative 21; Mismatches 57; Indels 47; Gaps 8;
QY 20 IHYDKITEEINKAIDD---AIAAEQSETIDPMKVPDHADKPERHVGIVDF----- 67
DB 478 IRYDNSQVIAAILDNLPLDAQMDRQEVYIP---PDQDKQQRQTGLRHFVHGDY 534
QY 68 -----KG-ELAMRNIE-----ARGLKQMKRQGDANVKGEIGVKAHLIG 106
DB 535 DVLTRDQPECIKQKGLPGAPRNAEQLLDDKEDLKQKERYQKYAMVEETTLER----G 590
QY 107 VHDIVSMEDYLAYL---GDLHPTTHVISDIOFVVALSLEISDE 149
DB 591 EYDD---EYDYSVEAQNEGOAPPVSLRGLQEAASSAYNAQDE 632

RESULT 99

F89819
endopeptidase [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 17-May-2002
C:Accession: F89819
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mitutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89819
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-818 <KUR>
A:Cross-references: GB:BA000018; PID:gl3700415; PIDN:BAB41713.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: clpC
C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 7.4%; Score 78.5; DB 2; Length 818;
Best Local Similarity 22.6%; Pred. No. 2.6e+02;
Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;
QY 18 DPIHYDKITEE-----INKAIDDAIAAEQSETIDPMKVPDHADKPERHVGIVDFKGE 72
DB 504 DTLHERVIGQKDAVNSISKAVRRARAGIK-----DP-----KRPIGSFILGPTG 548
QY 73 MENIE-ARGL-KQMKRQGDANVKGE-EGIVKAHL---LIGV-----HDDIVSMEDYLA 120
DB 549 VKTELARALAESMFGDDAMIRVDMSEFMKHAIVSLVGAPPYVGHDDGGQTEKVR 608
QY 121 KLGDLHPTTHVISD-----IQDFVVALSLEISDEGNITMTSFEVQFANVV-----NH 172
DB 609 K-----PYSVILFDEIEKAHPDVENILLQVLDDGHLTDTKGRTVDFNTIIIMTSNVA 663
QY 173 SILDPIFGVLSDLTAIFQDTRVKEMTKVLAPAKRE 209
DB 664 ELQDQRFAGFGGSGDQDYETIRKTMKLNKNSRPE 700

RESULT 100

E81681
proteinase, Lon family TC0623 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2003
C:Accession: E81681
R;Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.P.; White, O.; Hickey, I.; C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: E81681
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <RET>
A:Cross-references: GB:AE002331; GB:AE002160; NID:g7190662; PIDN:AAF39454.1; PID:g7190662
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0623
C:Superfamily: ATP-dependent Lon protease

Query Match 7.4%; Score 78.5; DB 2; Length 819;
Best Local Similarity 21.8%; Pred. No. 2.6e+02;
Matches 51; Conservative 38; Mismatches 78; Indels 67; Gaps 11;
QY 16 SADPIHYDKITEEINKAIDDAIAAEQSETIDP-----MKVP----- 52
DB 6 NTDSQNLDPNASEVEKELIDESAEEKTDHPTPSELFILPNKRPFFPGMAAPLLIEAG 65
QY 53 DH-----ADKFERHVGIVDFPKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLI 105
DB 66 PHVEVLTLLAKSSQKHIGIVLTKEKA--NTLKIGFNQLHRVG-----VSARIL- 112
QY 106 GVHDDIVSMEDYLAYLKGLDHPHTTHVISDIOFVVALSLEISDEGNITMTSFEVQFAN 164
DB 113 -----RIMPEGGSAQVLLSIEDRIIVKPVQKYLKAKVAHYKE-NKELTE-ELKAYSIS 166
QY 165 VYNNHIGGLSILDPIF-----GVLSDLTAIFQDTRVKEMTKVL 202
DB 167 IVSIIKDLKLNPLFKKEIQIFLGHSDFTPEPKGLADFSVALTTAT-RBELQEV 219

Search completed: August 6, 2004, 16:02:10
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 15:54:41 ; Search time 14 Seconds
(without alignments)
792.210 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068

Sequence: 1 MMKFLIIAAVAVVAVSADPI.....VRKEMTKVLAPAKRELEKN 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1068	100.0	213	1 ALL7_DERFA	Q26456 dermatophag
2	962	90.1	215	1 ALL7_DERPT	P49273 dermatophag
3	250.5	23.5	216	1 ALL7_LEPDS	Q91922 lepidoglyph
4	103	9.6	436	1 ENGA_STRMU	O8d90 streptococc
5	91.5	8.6	1393	1 RPOC_CHLPN	O92999 chlamydia p
6	90.5	8.5	547	1 CH60_XYLFA	O9pf02 xylella fas
7	90.5	8.5	547	1 CH60_XYLFT	O87bc0 xylella fas
8	90.5	8.5	547	1 CH61_VIBPA	Q91795 vibrio para
9	89.5	8.4	1539	1 Y373_HUMAN	O15078 homo sapien
10	88.5	8.3	458	1 FKX4_HUMAN	O02V90 homo sapien
11	88.5	8.3	1453	1 Y373_BOVIN	O9tu23 bos taurus
12	88	8.2	268	1 SPED_BUCAP	O8k9t6 buchnera ap
13	88	8.2	436	1 ENGA_STRPN	O97pc9 streptococc
14	88	8.2	547	1 CH60_LEGPN	P26878 legionella
15	87.5	8.2	1011	1 SCA4_RICMN	Q9aj62 rickettsia
16	87	8.1	436	1 ENGA_STRP3	O8k18 streptococc
17	87	8.1	436	1 ENGA_STRPY	O9aid0 streptococc
18	86.5	8.1	457	1 FKX4_MOUSE	P30416 mus musculus
19	86.5	8.1	544	1 CH61_VIBCH	O9km77 vibrio chol
20	85.5	8.0	553	1 CH60_BUCTH	O8kx1 buchnera ap
21	85.5	8.0	903	1 YB56_METJA	Q58556 methanococc
22	85.5	8.0	1396	1 RPOC_CHLMU	O9pk79 chlamydia m
23	85.5	8.0	1396	1 RPOC_CHLWU	O84316 chlamydia t
24	85	8.0	550	1 CH60_BUCTS	O8kix2 buchnera ap
25	84.5	7.9	539	1 CH60_SERRU	O66202 serratia ru
26	84.5	7.9	540	1 CH60_KLFOR	O66214 klebsiella
27	84.5	7.9	540	1 CH60_KLBOX	O66210 klebsiella
28	84.5	7.9	1887	1 RPBL_DROME	P04052 drosophila
29	84	7.9	358	1 SYFA_STRCO	O88055 streptomyc
30	84	7.9	392	1 DEOB_STAAM	O99x76 staphylococ
31	84	7.9	513	1 XYLG_ECOLI	P37388 escherichia
32	83.5	7.8	540	1 CH60_ENTGE	O66194 enterobacte
33	83.5	7.8	546	1 CH61_VIBVU	Q9ala9 vibrio vuln

RESULT 1

ALIGNMENTS

34	83.5	7.8	548	1	CH60_YERPE
35	83.5	7.8	550	1	CH60_YEREN
36	83	7.8	244	1	R34E_METJA
37	83	7.8	465	1	APEA_CIOAB
38	83	7.8	518	1	PUR9_SYNEL
39	83	7.8	543	1	CH60_BACFO
40	83	7.8	633	1	DNAK_AGRF5
41	83	7.8	1060	1	DF3A_LACIA
42	82.5	7.7	279	1	CBIO_METJA
43	82.5	7.7	539	1	CH60_ENTAM
44	82.5	7.7	540	1	CH60_KLEPL
45	82.5	7.7	540	1	CH60_SERFI
46	82.5	7.7	546	1	CH60_XANAC
47	82.5	7.7	2483	1	COA2_HUMAN
48	82	7.7	461	1	DLDH_CHLPN
49	82	7.7	2104	1	MYS3_SCHPO
50	81.5	7.6	541	1	CH60_PANAN
51	81.5	7.6	547	1	CH60_HAEDU
52	81.5	7.6	548	1	CH61_VIBHA
53	81.5	7.6	745	1	ATCS_SYNY3
54	81	7.6	244	1	RS4E_METVA
55	81	7.6	445	1	GID_AQUAE
56	81	7.6	798	1	UNR_HUMAN
57	80.5	7.5	159	1	BFR_SERMA
58	80.5	7.5	194	1	VAT8_SULTO
59	80.5	7.5	539	1	CH60_ENTAG
60	80.5	7.5	539	1	CH60_ENTAS
61	80.5	7.5	540	1	CH60_ENTAB
62	80.5	7.5	540	1	CH60_ERWAP
63	80.5	7.5	541	1	CH60_ERWHE
64	80.5	7.5	1010	1	SCA4_RICPA
65	80	7.5	160	1	TPC2_CABEL
66	80	7.5	590	1	SYFB_METXA
67	80	7.5	798	1	UNR_RAT
68	80	7.5	1816	1	LMA4_MOUSE
69	79.5	7.4	457	1	FXB4_RABIT
70	79.5	7.4	540	1	CH60_ERWCA
71	79.5	7.4	547	1	CH60_SALTY
72	79.5	7.4	819	1	LOH_CHLTR
73	79.5	7.4	906	1	CTN2_CHICK
74	79	7.4	153	1	AKRG_CLOTE
75	79	7.4	220	1	KPYK_THELI
76	79	7.4	406	1	RL4B_ARATH
77	79	7.4	486	1	LE11_PYRAB
78	79	7.4	867	1	SYA_FUSNN
79	78.5	7.4	326	1	LIPC_SOLTU
80	78.5	7.4	351	1	YB72_HAETN
81	78.5	7.4	505	1	LE12_METHR
82	78.5	7.4	528	1	YAC8_YEAST
83	78.5	7.4	532	1	CH62_VIBVU
84	78.5	7.4	546	1	CH60_XANCH
85	78.5	7.4	549	1	CH60_COLMA
86	78.5	7.4	819	1	LOH_CHLMU
87	78.5	7.4	892	1	SYA_METJA
88	78.5	7.4	1081	1	SVI_TETTH
89	78	7.3	328	1	SYFA_BUCBP
90	78	7.3	407	1	RL4A_ARATH
91	78	7.3	555	1	CH60_RICTS
92	78	7.3	664	1	Y4FB_RHISN
93	78	7.3	1012	1	POLS_IBDVS
94	78	7.3	1012	1	POLS_IBDVS
95	77.5	7.3	191	1	VATE_SULAC
96	77.5	7.3	458	1	RADA_BACSU
97	77.5	7.3	539	1	CH60_ENTIT
98	77.5	7.3	545	1	CH60_SODGL
99	77.5	7.3	546	1	CH60_ACTPL
100	77.5	7.3	546	1	CH60_XANCP

Q8ziy3	yersinia pe
P48219	yersinia en
P54039	methanococc
Q97k30	clostridium
Q8dins	s bifunctio
P81284	bacteroides
P50019	agrobacteri
Q9ci70	lactococcu
Q58488	methanococ
O66196	enterobacte
O66212	klebsiella
O66204	serratia fi
O89p21	xanthomonas
O00763	homo sapien
Q92773	chlamydia p
O41577	schizosacch
O66218	pantoea ana
P31294	haemophilus
Q83w18	vibrio harv
P73241	synechocyst
P40223	methanococc
O66913	aquifex aeo
O75534	homo sapien
O68935	serratia ma
Q971b8	sulfolobus
O66200	enterobacte
O66198	enterobacte
O66222	erwinia aph
O66216	erwinia her
Q9aj75	rickettsia
O09665	caenorhabdi
Q8tx56	methanopyru
P13395	rattus norv
P79727	mus musculu
P27124	oryctolagus
O66220	erwinia car
P48217	salmonella
O84348	chlamydia t
P30997	gallus gall
Q894h3	clostridium
O8301	thermococcu
Q98f40	arabidopsis
Q9uz08	pyrococcus
O8xfj8	fusobacteri
P80471	solanum tub
P44206	haemophilus
O27525	methanobact
P39734	saccharomyc
Q8cwj0	vibrio vuln
Q8xit7	xanthomonas
Q93gt8	colwellia m
Q9pk50	chlamydia m
O57984	methanococc
P36422	tetrachyena
P59504	buchnera ap
P49621	arabidopsis
P16625	rickettsia
P55440	rhizobium s
P25219	avian infec
P2351	avian infec
P22722	sulfolobus
P37572	bacillus su
O66192	enterobacte
Q9anr8	sodalis glo
P94166	actinobacil
Q8pd23	xanthomonas

ALL7 DERFA
ID ALL7_DERFA STANDARD; PRT; 213 AA.
AC Q26456;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mite allergen Der f 7 precursor (Der f VII).
GN DERF7.
OS Dermatophagoides farinae (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Psoroptidia; Analgoidea;
OC Pyroglyphidae; Dermatophagoides.
OX NCBI_TaxID=6954;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120794; PubMed=8556554;
RA Shen H.-D., Chua K.-Y., Lin W.L., Hsieh K.-H., Thomas W.R.;
RT "Molecular cloning and immunological characterization of the house
dust mite allergen Der f 7."
RL Clin. Exp. Allergy 25:1000-1006(1995).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALLERGEN: Causes an allergic reaction in human.
CC -1- SIMILARITY: Belongs to the mite group 7 allergen family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; S80655; AAB35977.1; -.
KW Allergen; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 213 MITE ALLERGEN DER F 7. (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC...). (POTENTIAL).
SQ SEQUENCE 213 AA; 23627 MW; 3CF1F529107B7808 CRC64;
Query Match 100.0%; Score 1068; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 3.4e-79;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDKFER 60
DB 1 MMKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDKFER 60
QY 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMYDLY 120
DB 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMYDLY 120
QY 121 KGLDLHPHTHVISIDQDFVVALSLEISDEGNITMTSFEVROFANVNHIGLSILDPIFG 180
DB 121 KGLDLHPHTHVISIDQDFVVALSLEISDEGNITMTSFEVROFANVNHIGLSILDPIFG 180
QY 181 VLSDLVLTAFQDVTVRKENTKVLAPAFKRELEKN 213
DB 181 VLSDLVLTAFQDVTVRKENTKVLAPAFKRELEKN 213
RESULT 2
ALL7 DERPT
ID ALL7_DERPT STANDARD; PRT; 215 AA.
AC P49773.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mite allergen Der p 7 precursor (Der p VII).
GN DERP7.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Psoroptidia; Analgoidea;
OC Pyroglyphidae; Dermatophagoides.

OX NCBI_TaxID=6956;
RN [1]
RP SEQUENCE FROM N.A.
RX Shen H.-D., Chua K.
RT "Molecular cloning
binding specificit;
Clin. Exp. Allergy
CC -1- SUBCELLULAR LO
CC -1- ALLERGEN: Caus
CC -1- SIMILARITY: Be
CC This SWISS-PROT en
CC between the Swiss
CC the European Bioinformatics Institute
CC use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; U37044; AAB80264.1; -.
KW Allergen; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 215 MITE ALLERGEN DER P 7. (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC...). (POTENTIAL).
SQ SEQUENCE 215 AA; 23881 MW; 63AD03DB5C8B1C0 CRC64;
Query Match 90.1%; Score 962; DB 1; Length 215;
Best Local Similarity 85.9%; Pred. No. 1.2e-70;
Matches 183; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
QY 1 MMKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDKFER 60
DB 1 MMKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDKFER 60
QY 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMYDLY 120
DB 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMYDLY 120
QY 121 KGLDLHPHTHVISIDQDFVVALSLEISDEGNITMTSFEVROFANVNHIGLSILDPIFG 180
DB 121 KGLDLHPHTHVISIDQDFVVALSLEISDEGNITMTSFEVROFANVNHIGLSILDPIFG 180
QY 181 VLSDLVLTAFQDVTVRKENTKVLAPAFKRELEKN 213
DB 181 VLSDLVLTAFQDVTVRKENTKVLAPAFKRELEKN 213
RESULT 3
ALL7 LEPDS
ID ALL7_LEPDS STANDARD; PRT; 216 AA.
AC Q9UG2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mite allergen Lep d 7 precursor.
OS Lepidoglyphus destructor (Storage mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Glycyphagoidea; Glycyphagidae;
OC Lepidoglyphus.
OX NCBI_TaxID=36936;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21099292; PubMed=11168362;
RA Eriksson T.L.J., Rasool O., Huecas S., Whitley P., Cramer R.,
RA Appenzeller U., Gafvelin G., van Hage-Hanstén M.;
RT "Cloning of three new allergens from the dust mite Lepidoglyphus
destructor using phase surface display technology."
RL Eur. J. Biochem. 268:287-294(2001).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALLERGEN: Causes an allergic reaction in human.
CC -1- SIMILARITY: Belongs to the mite group 7 allergen family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Baas S.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bowan R.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Winn M., Nelson W., DeBoy R., Koonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CML029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW-183;
 RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
 RT other Chlamydia strains based on whole genome sequence analysis.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 CC of DNA into RNA using the four ribonucleoside triphosphates as
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: The enzyme consists of the sigma chain and the core
 CC enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
 CC beta' chain.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 DR EMBL; AE001593; AAD18235.1; -;
 DR EMBL; AE002228; AAF38501.1; ALT_INIT.
 DR EMBL; AF002545; BAA98292.1; -;
 DR EMBL; AE017157; AAP98015.1; -;
 DR PIR; B86501; B86501.
 DR PIR; E72122; E72122.
 DR HSSP; O8KXU6; 1HQW.
 DR TIGR; CP0693; -;
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR007080; RNA_pol_Rpb1_1.
 DR InterPro; IPR007066; RNA_pol_Rpb1_3.
 DR InterPro; IPR007083; RNA_pol_Rpb1_4.
 DR InterPro; IPR007081; RNA_pol_Rpb1_5.
 DR InterPro; IPR006592; RNA_pol_N.
 DR Pfam; PF04997; RNA_pol_Rpb1_1.
 DR Pfam; PF06623; RNA_pol_Rpb1_2.
 DR Pfam; PF04983; RNA_pol_Rpb1_3.
 DR Pfam; PF05000; RNA_pol_Rpb1_4.
 DR Pfam; PF04998; RNA_pol_Rpb1_5.
 DR SMART; SM00663; RPOA_N.1.
 KW Transferase: DNA-directed RNA polymerase; Transcription;
 KW Complete proteome. 1031
 FT CONFLICT 1031 A -> G (IN REF. 1).
 SQ SEQUENCE 1393 AA; 154900 MW; E0734EF236C6FDE8 CRC64;

Query Match
 Best Local Similarity 21.2%; Pred. No. 16;
 Matches 53; Conservative 35; Mismatches 63; Indels 99; Gaps 12;

QY 36 AIAAIEQSETIDP-----MKYPDHADKFERHVGYD 66
 DB 1106 AIISVEGQVRDPGMLLAELPRGAITKXDTGSLPRVAELVEARPEADADAKIDGVVD 1165
 QY 67 FKGEIARMI-----EARGLKQ-----MKRQGDANVKE---EGIVKAHLI--- 105
 DB 1166 FKGIQNKRIILVVCDMDTGMEEHLIPLTKLIVQGDGVKQQLTDLGLVVPHEILETC 1225
 QY 106 GYHD--DIVSMEDLAYKGLDLPHTHVISDIDQFVAL-----SLEISDEGNITWTS 156
 DB 1226 GYRELQKYLNEVQEVYRLQGV-----DINKHIEIIVQMLQKVRITDPGDTILL- 1276
 QY 157 FEVRQFANVYN-----HIGG-----LSILDPIFGLVSDVITAFQDVTVR 195
 DB 1277 -----FGEDVNKFEYENRTEEDGGKPAQVPLLGITKASLTGTFESIAASFQDT-- 1329
 QY 196 KEMTKVLAPA 205
 DB 1330 ---TRVLTD 1336
 RESULT 6
 CH60_XYLFA
 ID CH60_XYLFA STANDARD; PRT; 547 AA.
 AC Q9PFF2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
 GN GROEL OR GROEL OR XF0615.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5G;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Buono M.R.P., Bonaccorsi A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Coutinho N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Mend C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
 CC proper assembly of unfolded polypeptides generated under stress
 CC conditions (By similarity).
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 CC 7 subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Roak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young J.W., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-24, SUBUNIT, AND FUNCTION.
RC TISSUE=Thymus;
RX MEDLINE=9285692; PubMed=1376003;
RA Tai P.-K.K., Albers M.W., Chang H., Faber L.E., Schreiber S.L.;
RT "Association of a 59-kilodalton immunophilin with the glucocorticoid
RT receptor complex.";
RL Science 256:1315-1318(1992).
RN [4]
RP SEQUENCE OF 1-20, FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RC TISSUE=Lymphocytes;
RX MEDLINE=9335211; PubMed=2378870;
RA Sanchez E.R., Faber L.E., Henzel W.J., Pratt W.B.;
RT "The 56-59-kilodalton protein identified in untransformed steroid
RT receptor complexes is a unique protein that exists in cytosol in a
RT complex with both the 70- and 90-kilodalton heat shock proteins.";
RL Biochemistry 29:5145-5152(1990).
RN [5]
RP SEQUENCE OF 1-17.
RC TISSUE=T-cell;
RX MEDLINE=92147620; PubMed=1371107;
RA Yem A.W., Tomaselli A.G., Heinrikson R.L., Zurcher-Neely H.,
RA Ruff V.A., Johnson R.A., Deibel M.R. Jr.;
RT "The Hsp36 component of steroid receptor complexes binds to
RT immobilized FK506 and shows homology to FKBP-12 and FKBP-13.";
RL J. Biol. Chem. 267:2868-2871(1992).
CC -!- FUNCTION: Component of unactivated mammalian steroid receptor
CC complexes that sediment at 8-10 S. May have a rotanase activity.
CC May play a role in the intracellular trafficking of hetero-
CC oligomeric forms of steroid hormone receptors.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SUBUNIT: Interacts with NR3C1 and dynein (By similarity).
CC Associates with HSP90 and HSP70 in unactivated steroid hormone
CC receptor complexes. Also interacts with peroxisomal phytanoyl-CoA
CC alpha-hydroxylase (PHVH).
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- TISSUE SPECIFICITY: Widely expressed.
CC -!- PTM: Phosphorylation by CK2 results in loss of HSP90 binding
CC activity (By similarity).
CC -!- SIMILARITY: Belongs to the FKBP-type PPIase family.
CC -!- SIMILARITY: Contains 2 FKBP-type PPIase domains.
CC -!- SIMILARITY: Contains 3 TPR repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88279; AAA36111.1; -;
CC EMBL; BC001786; AAA01786.1; -;

DR EMBL; BC002887; AAA02887.1; -;
DR EMBL; BC007924; AAA07924.1; -;
DR PIR; A46372; A46372.
DR PDB; 1NIA; 30-DEC-02.
DR Genew; HGNC:3720; FKBP4.
DR MIM; 600611; -;
DR GO; GO:0005737; Cytoplasm; TAS.
DR GO; GO:0005528; F:FK506 binding; TAS.
DR GO; GO:0008077; P:Hsp70/Hsp90 organizing protein activity; TAS.
DR GO; GO:0006457; P:protein folding; TAS.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR008941; TPR-Like.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00254; FKBP; 2.
DR SMART; SM00515; TPR; 3.
DR SMART; SM00028; TPR; 3.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
DR PROSITE; PS00454; FKBP_PPIASE_2; 2.
DR PROSITE; PS00509; FKBP_PPIASE_3; 2.
DR Isomerase; Rotamase; Repeat; TPR repeat; Nuclear protein;
KW Phosphorylation; 3D-structure.
FT INIT MET 0
FT DOMAIN 49 137 PPIASE, FKBP-TYPE 1.
FT DOMAIN 166 252 PPIASE, FKBP-TYPE 2.
FT REPEAT 269 302 TPR 1.
FT REPEAT 318 351 TPR 2.
FT REPEAT 353 385 TPR 3.
FT MOD_RES 142 142 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT CONFLICT 145 146 ED -> AR (IN REF. 2; AAA02887).
SQ SEQUENCE 458 AA; 51673 MW; 916B3B945C51634E CRC64;
Query Match 8.3%; Score 88.5; DB 1; Length 458;
Best Local Similarity 24.6%; Pred. No. 7.5; Indels 17; Gaps 3;
Matches 30; Conservative 25; Mismatches 50;
QY 18 DPHVDKITEINKAIDDAIAAIEQSETI-----DFMKVPDHD-KFERHV 62
Db 74 DKFSPDLGKGVKAWDIAIAIMKGVCHITCKPEYAYGAGSPKIPPNATLVFE--V 131
QY 63 GIVDFKGLAMNTEARGCLKQKQGDANVKGEGIVKHAHLIGVHDDIVSMYDLYKL 122
Db 132 ELFFKFGEDLTTEEDGGIIRIQTRGEGYAKPNEGAIIVEVALEGYKDKLPDQRELPEI 191
QY 123 GD 124
Db 192 GE 193
RESULT 11
ID Y373 BOVIN STANDARD; PRT; 1453 AA.
AC Q9TU23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RA Jovov B., Ripoll P.J., Benos D.J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88279; AAA36111.1; -;
CC EMBL; BC001786; AAA01786.1; -;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR: EMBL; AF176816; AAF00990.1; -
 CC KW Hypothetical protein; Coiled coil.
 FT DOMAIN 37 1426 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF139 CRC64;

Query Match 8.3%; Score 88.5; DB 1; Length 1453;
 Best Local Similarity 17.5%; Pred. No. 29;
 Matches 27; Conservative 43; Mismatches 69; Indels 15; Gaps 3;

QY 3 KFLIIAFAVAVSADPHYKITEINKAIDDA---IAAIEOSE---TIDPMKVPDHA 55
 DB 179 KPAELTRINLEAQAQVEQMLRDELADSVKTSVDADROHILELSEKEMELKVEVSKLEIS 238
 QY 56 DKFERHVGIVDFKGLAMRNIEARGLKQKRGDANVKRGEGIVKHAHLIGVHDDIVSME 115
 DB 239 DIAKQVEILNAQQOSREKEVESLRTQLLDYCAQSDKA-----LIAKLHQHVVSQ 290
 QY 116 YDLAVKGLDLPHTTHVSDIQDFVVALSLEISDE 149
 DB 291 ASEAALGKVESVASKLQKVEAHTLRLEQKLDEK 324

RESULT 12

SPED_BUCAP STANDARD; PRT; 268 AA.
 AC Q8K9T6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdoMetDC)
 DE (SamDC) [contains: S-adenosylmethionine decarboxylase beta chain; S-adenosylmethionine decarboxylase alpha chain].
 GN SPED OR BUSG202.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=98794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22084549; PubMed=12089438;
 RA Tamas I., Klaassen L., Canbaeck B., Naeslund A.K., Eriksson A.-S., Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";
 RL Science 296:2376-2379(2002).
 CC -!- FUNCTION: Decarboxylation of S-adenosylmethionine provides the aminopropyl moiety required for spermidine biosynthesis from putrescine (By similarity).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine = (5-deoxy-5-adenosyl) (3-aminopropyl) methylsulfonium salt + CO(2).
 CC -!- COFACTOR: Pyruvoyl group (By similarity).
 CC -!- SIMILARITY: Belongs to the prokaryotic AdoMetDC family. Subfamily 2.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AE014096; AAM67766.1; -
 DR HAMAP; MF_00465; -; 1.
 DR InterPro; IPR003826; SAMDC.
 DR Pfam; PF02675; AdoMetDC; 1.
 DR PIRSF; PIRSF001356; SAM decarboxylase; 1.
 KW Spermidine biosynthesis; Lyase; Decarboxylase; Zymogen; Pyruvate;
 FT CHAIN 1 112 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA CHAIN (BY SIMILARITY).

FT CHAIN 113 268 S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA CHAIN (BY SIMILARITY).
 FT SITE 112 113 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).
 FT MOD_RES 113 113 CONVERTED TO A PYRUVOYL GROUP (BY SIMILARITY).
 SQ SEQUENCE 268 AA; 31252 MW; 69A5F94715D2CC57 CRC64;

Query Match 8.2%; Score 88; DB 1; Length 268;
 Best Local Similarity 21.7%; Pred. No. 4.4;
 Matches 47; Conservative 38; Mismatches 64; Indels 68; Gaps 11;

QY 13 VAVSADPHYDK--ITEINKAIDDAIAIEQSETIDPMKVPDHAKEF-----ERH- 61
 DB 80 ILVCEPINKLVNKSNKIISSVLA-----HLDKSHCVHTVPESH 125
 QY 62 -VGIVDFKGLAMRNIEARGLKQKRGDANVKRGEGIVKA-----HLLIGHDDIVSME 115
 DB 126 QSGICTFRADIEVSTC-----GIISPLNALNYLIHQLESIVITIE 165
 QY 116 YDLAVKGLDLPHTTHV---ISDIQDFV---VALSLEISD---EGNITMTSFVEVQFAN 164
 DB 166 YVRGFTRIDHIGKHPIDHKINSIQNFMSHDIKSMYEMFDVNIYQENIFHTRMLKEF-N 224
 QY 165 VVNHHGGISILDPFGVLSVLTALFQDVTVRKEMTKV 201
 DB 225 LKNYLFNINVKD-----LSKKEHSYIIDLLWKEMREI 256

RESULT 13

ENGA_STRPN STANDARD; PRT; 436 AA.
 ID ID_ENGA_STRPN
 AC Q97PC9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE GTP-binding protein enga.
 GN ENGA OR SERA OR SP1709 OR SPRI553.
 OS Streptococcus pneumoniae, and
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313, 171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D., Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., E.K., McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
 RL Science 293:498-506(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-255 / R6;
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burtett S., DeHoff B.S., Bateman S.T., Fritz L., Fu D.-J., Fuller W., Geringer C., Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E., LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P., McAdhrien S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I., Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P., Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G., Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L., Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717(2001)
 CC -!- FUNCTION: GTPase of unknown physiological role.


```
AC Q9AJB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment)
GN SCA4 OR D.
OS Rickettsia mongolotimonae.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=45261;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT 'Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL "gene D" coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF151725; AAK30683.1; -
KW Antigen.
FT NON_TER 1
FT NON_TER 1011
FT NON_TER 1011
SQ SEQUENCE 1011 AA; 110607 MW; 0169A06981BDD5D08 CRC64;

Query Match 8.2%; Score 87.5; DB 1; Length 1011;
Best Local Similarity 22.4%; Pred. NO. 23;
Matches 55; Conservative 35; Mismatches 98; Indels 57; Gaps 10;

QY 5 LLIAAVAFVAVSADEPIHYDK--ITEEI--NKAID--DATAIEQSETIDPMKVPDHAQK 57
DB 402 LLNTATALSGSMQDLNLYNAGLTKEIDSNKQIDLIKEAATAILNNE-----KSDI 452
QY 58 FERHVGIVDFKGLAMRNIIEARGLKQMKRQGDANVKGEEGIV-----KAHL 104
DB 453 AEKQANII-----ALAEIVNNKLP-----DAKVGAVNVALETIKNDQTPNLEKML 503
QY 105 IGVEDDIVSME-----YDLAYKGLDHPHTTHVISDIQDFVVALSLEISDEG 150
DB 504 EATVAIILNSENLEFPKQKQOMLEKXAVDVGLSLKDDASRAVTIDGIDKDWIKTNLSTEDKG 563
QY 151 NITWT--SPFVROFANVNVHIGLSILDPFGVLSVLTAIFQDVTVRKEMTKVLAPAFK 207
DB 564 TMLTAGVDKXNVSELNNAEKSKSLGSLVKK--GVEAQVLSPAQQLIQHLDKITAEOIK 621
QY 208 RELEK 212
DB 622 KDTIK 626

RESULT 16
ID ENGA STRP3 STANDARD; PRT; 436 AA.
AC Q8K8J6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE GTP-binding protein engA.
GN ENGA OR PGDA OR SPYM3_0249 OR SP51610.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345;
RA Okagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
CC -!- FUNCTION: GTPase of unknown physiological role.
CC -!- SIMILARITY: Belongs to the era/trmE family of GTP-binding
CC proteins. EngA subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014141; RAU78856.1; -
DR EMBL; AP005146; BAC64705.1; -
DR HAMAP; MF_00195; -; 1.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR006073; GTP_OBG.
DR InterPro; IPR002917; MMR_HSR1.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01926; MMR_HSR1; 1.
DR PRINTS; PR00326; GTP_OBG.
DR TIGRPFAMs; TIGR00650; MG442; 2.
DR TIGRPFAMs; TIGR00231; small_GTP; 2.
KW GTP-binding; Repeat; Complete proteome.
FT NP_BIND 10 17 GTP 1 (POTENTIAL).
FT NP_BIND 57 61 GTP 1 (POTENTIAL).
FT NP_BIND 119 122 GTP 1 (POTENTIAL).
FT NP_BIND 181 188 GTP 2 (POTENTIAL).
FT NP_BIND 229 233 GTP 2 (POTENTIAL).
FT NP_BIND 294 297 GTP 2 (POTENTIAL).
SQ SEQUENCE 436 AA; 48771 MW; F2DF6CD5B674323A CRC64;

Query Match 8.1%; Score 87; DB 1; Length 436;
Best Local Similarity 22.2%; Pred. NO. 9.4;
Matches 47; Conservative 41; Mismatches 80; Indels 44; Gaps 11;

QY 30 NKAIDDAIAAQSETIDPMKVPDHAQKFERHVGIVDFKGLAMRNIIEARGLKQMKRQG- 88
DB 21 NRIAGERISIVDEGVGTRDIYATGEWLNQFSLDTGG---IDDDVAPNEQIKHQAG 77
QY 89 -----DAN-----VKGEEGIVKAH-----LLIGVHD-DIVSMEYDL-----AYK 121
DB 78 IAMEEADVIVFVSGKEGVTDADDEVSVSKILYRTNTPVILAVNKVDNPMRNDIVDFYSLG 137
QY 122 LGLDHPHTTHV-----LSDIQDFVVA--LSLEISDEGNITWTSEFVRQFANV-----VNHIGG 171
DB 138 LGDPYPVSSVHGIGTGDVDAIVENLPVEBAEE--NDDIRFSLICRPNVNGKSLINAILG 196
QY 172 LS--ILDPFGVLSVLTAIFQDVTVRKEMTKV 201
DB 197 EDRVIASPVAGTTRDAIDTFTDADGQEFMTI 228
```

```

RESULT 17
ENGA_STRPY STANDARD; PRT; 436 AA.
AC Q9ALDO;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE GTP-binding protein, ena.
GN ENGA OR F04A OR SPY0341 OR SPYM18_0334.
OS Streptococcus pyogenes, and
OC Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 186103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvoarov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar P.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barthian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.B., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
CC -1- SIMILARITY: GPase of unknown physiological role.
CC -1- SIMILARITY: Belongs to the era/time family of GTP-binding
CC proteins. ENGA subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006498; AKK33393.1; -
DR EMBL; AE009978; AAL97088.1; -
DR HAWAP; MF_00195; -; 1.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR006073; GTP1_OBG.
DR InterPro; IPR002917; MMR_HSR1.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01926; MMR_HSR1; 1.
DR PRINTS; PR00326; GTP_OBG.
DR TIGRFAMs; TIGR00650; MG442; 2.
DR TIGRFAMs; TIGR00231; small_GTP; 2.
KW GTP-binding; Repeat; Complete proteome.
FT NP_BIND 10 17 GTP 1 (POTENTIAL).
FT NP_BIND 57 61 GTP 1 (POTENTIAL).
FT NP_BIND 119 122 GTP 1 (POTENTIAL).
FT NP_BIND 181 188 GTP 2 (POTENTIAL).
FT NP_BIND 229 233 GTP 2 (POTENTIAL).
FT NP_BIND 294 297 GTP 2 (POTENTIAL).
SQ SEQUENCE 436 AA; 48801 MW; E9DED7D51675323A CRC64;

Query Match 8.1%; Score 87; DB 1; Length 436;
Best Local Similarity 22.2%; Pred No. 9.4;
Matches 47; Conservative 41; Mismatches 44; Gaps 11;

```

```

QY 30 NKAIDDAIAAEQSETIDPMKVPDHADKFERHVGIVDFKGLAMVNTIARGLKQKRG- 88
DB 21 NRIAGERISIVEDVEGVTRDRIYATGEWLNQFSLDTGG---IDDVDPMEQIKHQAG 77
QY 89 ----DAN-----VKGEIGIVKAH-----LLIGVHD-DIVSMEYDL-----AYK 121
DB 78 IAMEEADVIVFVSGKGVTDADDEVSVKILYRTNTFTVILAVNKVDNPEMRNDIVDFYSLG 137
QY 122 LGDLHPPTHV-----ISDIQDFVVA-LSLEISDEGNITWTSFEVRFQFANY-----VNHHGG 171
DB 138 LGDPYVSSVHGIGTGDVLDIAENLPLVEAAE-NDDIIRFSLIGRPNVGKSLINAILG 196
QY 172 LS--ILDPIFGVLSVLTALTAIFQDITVRKEMTKV 201
DB 197 EDRVIASPVAGTTRDAIDTHTDADGQEFMTI 228

RESULT 18
FKB4_MOUSE
ID FKB4_MOUSE STANDARD; PRT; 457 AA.
AC P30416;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE FK506-binding protein 4 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans
DE isomerase) (PPIase) (Rotamase) (p59 protein) (HSP binding
DE immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein)
DE (FKBP59).
DE FKB4.
GN FKB4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=94040772; PubMed=7693550;
RA Schmitt J., Stummenberg H.G.;
RT "Cloning and expression of a mouse cDNA encoding p59, an immunophilin
RT that associates with the glucocorticoid receptor.";
RL Gene 132:267-271 (1993).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Lung tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.N., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
SEQUENCE OF 5-457 FROM N.A., PARTIAL SEQUENCE, FUNCTION, AND
RP SUBCELLULAR LOCATION.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=93342084; PubMed=8341706;
RA Alnemri E.S., Fernandes-Alnemri T., Nelki D.S., Dudley K.,
RA Dubois G.C., Litwack G.;
RT "Overexpression, characterization, and purification of a recombinant
RT mouse immunophilin FKBP-52 and identification of an associated

```

phosphoprotein.";
 Proc. Natl. Acad. Sci. U.S.A. 90:6839-6843(1993).
 CC -!- FUNCTION: Component of unactivated mammalian steroid receptor
 CC complexes that sediment at 8-10 S. May have a rotamase activity.
 CC May play a role in the intracellular trafficking of hetero-
 CC oligomeric forms of steroid hormone receptors.
 CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
 CC (omega=0).
 CC -!- SUBUNIT: Associates with HSP90 and HSP70 in unactivated steroid
 CC hormone receptor complexes, also interacts with peroxisomal
 CC phytyl-CoA alpha-hydroxylase (PHYH). Interacts with NR3C1 and
 CC dyx1c1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
 CC -!- PTM: Phosphorylation by CK2 results in loss of HSP90 binding
 CC activity (By similarity).
 CC -!- SIMILARITY: Belongs to the FKBP-type PPIase family.
 CC -!- SIMILARITY: Contains 2 FKBP-type PPIase domains.
 CC -!- SIMILARITY: Contains 3 TPR repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X70887; CAA50231.1; -.
 CC EMBL; BC003447; AAH03447.1; -.
 CC EMBL; X17069; CAC39452.1; -.
 CC EMBL; X17068; CAA34914.1; ALT_SEQ.
 CC EIR; JN0873; JN0873.
 CC HSP; J27124; IROT.
 CC MGD; MGI:95543; Fkbp4.
 CC InterPro; IPR001179; FKBP_PPIase.
 CC InterPro; IPR008941; TPR-like.
 CC InterPro; IPR001440; TPR.
 CC Pfam; PF00254; FKBP; 2.
 CC Pfam; PF0515; TPR; 3.
 CC SMART; SM00028; TPR; 3.
 CC PROSITE; PS00453; FKBP_PPIASE 1; 1.
 CC PROSITE; PS00454; FKBP_PPIASE 2; 2.
 CC PROSITE; PS00509; FKBP_PPIASE 3; 2.
 CC IsoMase; Rotamase; Repeat; TPR repeat; Nuclear protein;
 CC Phosphorylation.
 CC INIT MET 0 BY SIMILARITY.
 CC DOMAIN 49 PPIASE, FKBP-TYPE 1.
 CC DOMAIN 166 PPIASE, FKBP-TYPE 2.
 CC REPEAT 269 TPR 1.
 CC REPEAT 318 TPR 2.
 CC REPEAT 352 TPR 3.
 CC MOD_RES 142 142.
 CC CONFLICT 5 6.
 CC CONFLICT 155 162.
 CC CONFLICT 202 239.
 CC
 CC CONFLICT 264 271.
 CC CONFLICT 314 314.
 CC SEQUENCE 457 AA; 51441 MW; 0D087C543541C1C0 CRC64;
 CC
 CC Query Match 8.1%; Score 86.5; DB 1; Length 457;
 CC Best Local Similarity 23.0%; Pred. No. 11;
 CC Matches 28; Conservative 27; Mismatches 50; Indels 17; Gaps 3;
 CC
 CC 18 DPVHYDKTEINKAIDDAIAIEQSETI-----DPMKVPDAD-KFERHV 62
 CC 74 DKFSDLKGGEVIRKWDIAVATMKVGEVCHTCKPEYAGNAGSPFKIPNATVFE--V 131
 CC
 CC 63 GIVDFPKGELAMNTEARGKVMKQGGDANVKGEGIVKVAHLLIGVHDDIVSMYVDLAYKL 122
 CC 132 ELFFPKGDLTEEDGGIIRIRTRGEGYARPNDGAMVEALEGYHKDRFLDQRELCEV 191

QY 123 GD 124
 DB 192 GE 193
 RESULT 19
 CH61_VIBCH STANDARD; PRT; 544 AA.
 ID CH61_VIBCH
 AC Q9KNR7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60 kDa chaperonin 1 (protein Cpn60 1) (groEL protein 1).
 GN GROEL1 OR GROEL1 OR VC2664.
 OS Vibrio cholerae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 CC NCBI_TaxID=666;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN=E1 Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
 CC proper assembly of unfolded polypeptides generated under stress
 CC conditions (By similarity).
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 CC 7 subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE004332; AAF95805.1; -.
 CC FIR; B82048; B82048.
 CC HSP; P06139; 1GRL.
 CC TIGR; VC2664; -.
 CC HAMAP; MF_00600; -. 1.
 CC InterPro; IPR001844; Chaperin Cpn60.
 CC InterPro; IPR002423; Cpn60/TCP-1.
 CC InterPro; IPR008950; GroEL-ATPase.
 CC Pfam; PF00118; Cpn60_TCP1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 CC PROSITE; PS00304; CHAPERONINS_CPN60; 1.
 CC Chaperone; ATP-binding; Complete proteome.
 CC SEQUENCE 544 AA; 57153 MW; 927868AE794AE079 CRC64;
 CC
 CC Query Match 8.1%; Score 86.5; DB 1; Length 544;
 CC Best Local Similarity 24.4%; Pred. No. 13;
 CC Matches 41; Conservative 21; Mismatches 69; Indels 37; Gaps 5;
 CC
 CC 6 LIAAFAVAVSRA-----DPIHYDKTEINKAIDDAIAIEQSETIDPMKV 51
 CC 372 LAGGVAVIKVGAATEVEMKEKDVEDDAHATRAAEEGVAGGVALLRAAKLSLV- 430
 CC 52 PDHAKDFERHVGIVDFKGLAMRNTEARGKVMKQGGDANVKGEGIVKVAHLLI 105
 CC 105

```
Db 431 ---GDNEQNGVI-----RVALRAMEAPLRQIVKVGADDEBSVANNVRAGEGNYGNAAT 482
Qy 106 GVHDDIVSMEDYLAAYKGLDLPHTTHVISDIQDFVVALSLEISDEGNIT 153
Db 483 GVGDMTEM-----GILDPKTVRSALQFAASVAGLMTTAMIT 522

RESULT 20
CH60_BUCTT STANDARD; PRT; 553 AA.
AC Q8KIXI;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL.
OS Buchnera aphidicola (subsp. Tuberolachnus salignus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2077640; PubMed=12082135;
RA Fares M.A., Barrio E., Sabater-Munoz B., Moya A.;
RT "The evolution of the heat-shock protein GroEL from Buchnera, the
RT primary endosymbiont of aphids, is governed by positive selection.";
RL Mol. Biol. Evol. 19;1162-1170(2002).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ439086; CAD27798.1; -.
DR HAMAP; MF 00600; -.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60_TCF1.1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 553 AA; 58010 MW; 814F300472676701 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 553;
Best Local Similarity 18.6%; Pred. No. 16;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

Qy 13 VASADPTHYDKITEINKAIDDA-IAAIEQSEITIDPMKVPDHADKFERHVGIVDFKGL 71
Db 107 VAAGMFW-----DLKRGIDKAVIAAVE-----LRKLSVPCSDSKAIAQVGTISANDL 156

Qy 72 AMRNIARGLKQMKRQGDANVKGEIGVKAHLIGVHDD---IVSMYDLYAY-----120
Db 157 TVGKLIAEAMGKV-----GKEGIVTVEEGTQLQDELDDVVEGQFQFGYLSPYFINK 207

Qy 121 -----KLGDLPHTTHVLSIDIQDFVVALSLEISDEGNITMTSFEV 159
Db 208 PESGALENPFILLADKKLSNRELLPFILEAVAKACKSLIIAEDVEGALATL-----262

Qy 160 RQFANVNHIGGL-----SILEPIFVGLSDVLTAFQDTVRKENTKVLAPAFKEELEK 212
```

```
Db 263 -----IVNTRMGIVKVAAPGFG---DRRKAMLDIATLTGGTVISEIGLELEK 311

RESULT 21
YB56_METJA STANDARD; PRT; 903 AA.
AC Q58556;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cell division cycle protein 48 homolog MJ1156.
GN MJ1156.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: Belongs to the AAA ATPase family. CDC48 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67557; AAS99153.1; -.
DR FIR; C64444; C64444.
DR TIGR; MJ1156; -.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003959; AAA ATPase_cent.
DR InterPro; IPR005938; AAA CDC48.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR009010; Asp decarb fold.
DR InterPro; IPR003338; ATPaseVAT_N.
DR InterPro; IPR004201; Cdc48_2.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02933; cdc48_2; 1.
DR Pfam; PF02359; cdc48_N; 1.
DR SMART; SM00382; AAA_2.
DR TIGRfam; TIGR01243; CDC48; 1.
DR PROSITE; PS00574; AAA; 2.
KW Hypothetical protein; ATP-binding; Repeat; Multigene family;
KW Complete proteome.
FT NP_BIND 220 227 ATP (POTENTIAL).
FT NP_BIND 493 500 ATP (POTENTIAL).
SQ SEQUENCE 903 AA; 100403 MW; 99F2ABEE2D544DC2 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 903;
Best Local Similarity 25.0%; Pred. No. 29;
Matches 53; Conservative 30; Mismatches 90; Indels 39; Gaps 11;

Qy 23 DKITEINKAIDD-----AIAIEQSEITIDPMKV-PDHADKFERHVGIVDFKGLAM 73
Db 571 DKVNNQLTELDGMEEPKDVVVIATNPDIIDPALLRPGRLDRVIL-VPVPDEKARLDI 629

Qy 74 RNIEARGLKQMKRQGDANV---KGEIGVKAHL-----LIGVHDDIVSMYDLYAY 121
```

Db 630 FKIHTR---SMNLAEVDNLELAKKTEGYTGADIEALCREAAMLAVERSI-GKPDWIEVK 695
 QY 122 LGDLHPTTHVTSIDQFVVALSLSEISDEGNIT--MTSPFVROQFANVNVNHHIGLSILDPIF 179
 Db 686 LREL---INYLQISGTFRAAAVELNSVTKATKESAEAGEFSELKNAIG-----KII 736
 QY 180 GVLSDVLTAFQDVTVRKEMTKVLAPAFKRELE 211
 Db 737 SVLSPAKEKI--EAVEKEIDKFLVEINKEELK 766

RESULT 22
 RPOC_CHLMU STANDARD; PRT; 1396 AA.
 AC Q9PK79;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
 DE beta' chain) (RNA polymerase beta' subunit).
 GN RPOC OR TC0588.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MoPn / Ni99;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 CC of DNA into RNA using the four ribonucleoside triphosphates as
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: The enzyme consists of the sigma chain and the core
 CC enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
 CC beta' chain.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF002327; AAF39420.1; --
 CC PIR; F81686; F81686.
 CC HSSP; Q9KWU6; 1HQW.
 CC TIGR; TC0588; --
 CC InterPro; IPR007022; RNA_pol_A.
 CC InterPro; IPR007080; RNA_pol_Rpb1_1.
 CC InterPro; IPR007066; RNA_pol_Rpb1_3.
 CC InterPro; IPR007083; RNA_pol_Rpb1_4.
 CC InterPro; IPR007081; RNA_pol_Rpb1_5.
 CC InterPro; IPR006592; RNA_pol_N.
 CC Pfam; PF04997; RNA_pol_Rpb1_1; 1.
 CC Pfam; PF06623; RNA_pol_Rpb1_2; 1.
 CC Pfam; PF04983; RNA_pol_Rpb1_3; 1.
 CC Pfam; PF05000; RNA_pol_Rpb1_4; 1.
 CC Pfam; PF04998; RNA_pol_Rpb1_5; 1.
 CC SMART; SM00663; RPOLA_N; 1.
 CC Transferrase; DNA-directed RNA polymerase; Transcription;
 KW Complete proteome.
 SQ SEQUENCE 1396 AA; 154897 MW; CC090B9A18986977 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 1396;
 Best Local Similarity 23.6%; Pred. No. 48;
 Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12;
 QY 43 STIDPMKVPDHDKFERHVGIVDFKGLAMENI-----EARGLKO-----MKRQ 87
 Db 1145 AELVEARKEPDAAD-IAKIDGVDFKQKQKRIILVVRDEVTEGMEELISLTKHLIVQR 1203
 QY 88 GDANVKGE---EGIVKAHLI---GVHD---DIVSMEYDLAYKGLDLPHTTHVISDIOFV 139
 Db 1204 GDSVIKQQLTDGLVVPHEILLEICGRELQKVLNNEVQEVYRLQGV-----DINDKH 1255
 QY 140 VAL-----SLEISDEGNITM---TSFEVRQFANV---VNHIGG-----LSILDLP 177
 Db 1256 IETIVRQMLQKVRITDPGDTLLFGEDVDKKEFYENRRTTEEDGKPAQAVPVLGITKA 1315
 QY 178 IFQVLSVLTALFQDVTVRKEMTKVLAPA 205
 Db 1316 SLGTESFISAAFSQDT-----TRVLTD 1338

RESULT 23
 RPOC_CHLTR STANDARD; PRT; 1396 AA.
 AC Q843I6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
 DE beta' chain) (RNA polymerase beta' subunit).
 GN RPOC OR CT314.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/Ck;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 CC of DNA into RNA using the four ribonucleoside triphosphates as
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: The enzyme consists of the sigma chain and the core
 CC enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
 CC beta' chain.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF001304; AAC67907.1; --
 CC PIR; G71529; G71529.
 CC HSSP; Q9KWU6; 1HQW.
 CC PHCI-2DPAGE; O84316; --
 CC InterPro; IPR007022; RNA_pol_A.
 CC InterPro; IPR007080; RNA_pol_Rpb1_1.
 CC InterPro; IPR007066; RNA_pol_Rpb1_3.
 CC InterPro; IPR007083; RNA_pol_Rpb1_4.
 CC InterPro; IPR007081; RNA_pol_Rpb1_5.
 CC InterPro; IPR006592; RNA_pol_N.

```

DR PRINTS; PR00296; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 550 AA; 1787 MW; 100D3B923DA48502 CRC64;

Query Match      8.0%; Score 85; DB 1; Length 550;
Best Local Similarity 21.7%; Pred. No. 18;
Matches 49; Conservative 42; Mismatches 87; Indels 48; Gaps 10;

Qy      13 VANSADPIHYDKITEINKAIDDA-IAAEQSTIDPMKVPDHDADPERHVGIVDPFGEL 71
      ||      ::      ::      ::      ::      ::      ::      ::      ::
Db      107 VAAGMNP-----DLKEGVDRKAVIAAEE-----LKCLSVPCSDFFKAIAGVGTISANSD 156

Qy      72 AMRNIERAGLIKMKROGDANVKEEGEIVKAHLILIGVHDD---TVSMEYDIAYLKLGLD--H 126
      ||      ::      ::      ::      ::      ::      ::      ::      ::
Db      157 TVGKLIAAEAKV-----GKEGVITVEEGTGLQDELDDVEGQDFRGYLSPYFINK 207

Qy      127 PTHIVISDIOQFVVALSLEISD-----GNITWISFEVROFA---NVVNHIG 170
      ||      ::      ::      ::      ::      ::      ::      ::      ::
Db      208 PDTGSELDNPFILLADKLLSNRMPLPLESVAKAGNGLLIIAEDVEGALATLVVNTMR 267

Qy      171 GL-----SILDPIFGVLSDLVLTAFQDFTVRKMTKVLAFAFKREK 212
      ::      ::      ::      ::      ::      ::      ::      ::
Db      268 GIWKVAAVKAPGFG---DRRKAMLODIATLTGSTVISEIGLEK 310

```

```

RESULT 25
CH60_SERRU STANDARD; PET; 539 AA.
ID CH60_SERRU
AC O66202;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROL OR GROEL OR WOPA.
OS Serratia rubideae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
NCBI_TaxID=61652;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=JCM 1240.
RC
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -I- FUNCTION: Prevents misfolding and promotes the refolding and
CC conditions (By similarity)
CC -I- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008143; BAA25219.1; -.
CC HSPSP; P06139; 1GRL.
CC HAMAP; MF 00600; -.
CC DR InterPro; IPR001844; Chaperin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR InterPro; IPR008950; GroEL-ATPase.
CC DR Pfam; PF00118; cpn60_TCP1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCP1.
CC

```

```

DR HAMAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-AtPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONING6.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR

```

```
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
FT NON_TER 540 540
SQ SEQUENCE 539 AA; 56640 MW; CB447B657515D4F0 CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 539;
Best Local Similarity 18.6%; Pred. No. 19;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPTHYDKITEINKAIDDA-IAAEQSEITIDPMKVPDHADKFERHVGIVDFKGL 71
DB 107 VAGGNWPM-----DLKRGIDKAVAAVEELKT---LSVPCSDSKAIAQVGTISANSD 156
QY 72 AMRNIARGLKMKRQGDANVKGEGIVKAHLIIQVHDD---IVSMYEDLAY----- 120
DB 157 TVGKLIAMKDV-----GKEGVITVEGTGLQDELVDVVEGMDRGLSPYPINK 207
QY 121 -----KLGDLHPTTHVISIDIQDFVVALSLEISDEGNITMTSFEV 159
DB 208 PETGSVELESPIILLADKKVSNIRELPVLEAVAKAGKPLIIVAEDEVGEALATL----- 262
QY 160 RQFANVNHIGL-----SILDPFGVLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATITAGTVISEEIGMELEK 311

RESULT 26
CH60_KLEOR STANDARD; PRT; 540 AA.
ID CH60_KLEOR STANDARD; PRT; 540 AA.
AC O66214;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOXA OR GROEL.
OS Klebsiella ornithinolytica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Raoultella.
OX NCBI_TaxID=54291;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 6096;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008149; BAA25231.1; -
CC HSSP; P06139; 1JON.
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00118; cpn60 TCP1; 1.
CC PRINTS; PR00296; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW
```



```

FT NON TER 540
SQ SEQUENCE 540 AA; 56445 MW; BFE441BC6B05E06B CRC64;

Query Match
Best local similarity 7.9%; Score 84.5; DB 1; Length 540;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPHYDKITEINKAIDA-TAAIFQSSTIDPMKVPDADKFEHVGIVDFKGL 71
DB 107 VAGMNPW-----DKRGIDKAVAAVEELK-----LSVFCSDSKAIAQVGISANSDE 156
QY 72 AMRNIARGKMKRQGDANVKGEEGIVKAHLIGVHDD---IVSMEYDIAY-----120
DB 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLEDELVDVEGQFDRGYLSPYFINK 207
QY 121 -----KLGDLHPHTHVLISDIQDFVALSLISISDEGNITWTSFEV 159
DB 208 PETGAVELESFILLADKISNIREMLFVLEAVAKAGKPLVIAEVEGEALATL-----262
QY 160 RQFANVNHIGGL-----SILDPIFGVLSVLTAFIOTVRKEMTKVLAFAFKRELEK 212
DB 263 -----VNVTRGVIKVAAVKAPGFG---DRRKAMLDQIATLTGGTVISSEIGNVELEK 311

RESULT 28
RPB1 DROME
ID RPB1 DROME STANDARD; PRT; 1887 AA.
AC P04052; Q9VYX6;
DT 01-NOV-1986 (Rel. 03, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).
GN RPII215 OR C01554.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89218930; PubMed=2496296;
RA Jokester R.S., Weeks J.R., Zehring W.A., Greenleaf A.L.;
RT "Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila."
RL Mol. Gen. Genet. 215:266-275 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster."
RN Science 287:2185-2195 (2000).
RP [3]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE=87089662; PubMed=3025586;
RA Seales L.L., Greenleaf A.L., Kemp W.E., Voelker R.A.;
RT "Sites of P element insertion and structures of P element deletions in the 5' region of Drosophila melanogaster RpiI215."
RL Mol. Cell. Biol. 6:3312-3319 (1986).
RN [4]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=85282618; PubMed=2992806;
RA Biggs J., Seales L.L., Greenleaf A.L.;
RT "Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II."
RL Cell 42:611-621 (1985).
RN [5]
RP SEQUENCE OF 1441-1887 FROM N.A.
RX MEDLINE=88094402; PubMed=3122024;
RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function."
RL Mol. Cell. Biol. 8:321-329 (1988).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).
CC -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO TWELVE DIFFERENT POLYPEPTIDES. THE 215 kDa POLYPEPTIDE IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: The tandem 7 residues repeats can be highly phosphorylated. The phosphorylation activates POL2.
CC -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases were found in eukaryotic nuclei: polymerase I for the ribosomal RNA precursor, polymerase II for the mRNA precursor, and polymerase III for 5S and tRNA genes.
CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27431; AAA28868.1; -.
DR EMBL; AE003486; AAP48057.1; -.
DR EMBL; M14203; AAA28864.1; -.
DR EMBL; M11798; AAA28863.1; -.
DR EMBL; M19537; AAA28827.1; -.
DR PIR; S04457; RNFF2L.
DR FlyBase; FBgn0003277; RpiI215.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.

```

```
DR InterPro; IPR007075; RNA_pol_Rpb1_6.
DR InterPro; IPR007073; RNA_pol_Rpb1_7.
DR InterPro; IPR006592; RNA_pol_Rpb1_N.
DR InterPro; IPR000684; RNA_polIT_repeat.
DR Pfam; PF04997; RNA_pol_Rpb1_1.
DR Pfam; PF00623; RNA_pol_Rpb1_2.
DR Pfam; PF04983; RNA_pol_Rpb1_3.
DR Pfam; PF05000; RNA_pol_Rpb1_4.
DR Pfam; PF04998; RNA_pol_Rpb1_5.
DR Pfam; PF04992; RNA_pol_Rpb1_6.
DR Pfam; PF04990; RNA_pol_Rpb1_7.
DR Pfam; PF05001; RNA_pol_Rpb1_R.
DR SMART; SM00663; RPOA1_N; 1.
DR PROSITE; PS00115; RNA_POL_II_REPEAT; 11.
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 67 83 C2H2-TYPE (POTENTIAL).
FT DOMAIN 1579 1881 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT CONFLICT 319 324 RAMQKS -> GYAKV (IN REF. 4).
FT CONFLICT 450 450 F -> G (IN REF. 4).
FT CONFLICT 455 458 TLHK -> RCTT (IN REF. 4).
FT CONFLICT 463 472 GHRVKVLPWS -> VTGESVASST (IN REF. 4).
FT CONFLICT 741 741 R -> H (IN REF. 1).
FT CONFLICT 1485 1524 SMLGGAAMFVGSGTSPMTPTWANCNTPRYFSPGHV
-> I (IN REF. 5).
FT CONFLICT 1506 1508 MTP -> ELDSA (IN REF. 1).
FT CONFLICT 1887 1887 D -> DVKKGGRG (IN REF. 1).
FT SEQUENCE 1887 AA; 209167 MW; 4EC68C7708A167A3 CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 1887;
Best Local Similarity 22.7%; Pred. No. 82;
Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;

QY 17 ADPIHYDKITEINKAIDDAIAEQSETIDPMKVPDHA--DKFERHVGIV-----DFKG 69
DB 686 ADPQTYNEIQAIKAKDDVINIQAHNMELETPGNTLRQTFENVKNVILNDARDKTG 745
QY 70 ELAMENI-EARGLKKMKRQ-----GDANKGEE--GIVKAHLILIGVHD 109
DB 746 GSAKSLTEYNNLKAMVYSGSGKSNINISQIVACGQGVNVEGKIPYGRKTLPHPIKD 805
QY 110 D-----IVSMEVDYLYKGLDLHPTT---HVISDIQDFVVALSLEISDSEGNI-----T 133
DB 806 DYGPESRGFVENSY-----LAGLTPSEFYFHAMGG-REGLIDTAVKTAETGYIQRLLIKA 859
QY 154 MTSFEVROFANVNHIGLSLIL---DPIFGVLSL----- 184
DB 860 MESVMNVYDGVTRNSVGGLIQIRYGEDGLGELVEFQNPVTKLSNKSFEKRFKFDWSNE 919
QY 185 -VLTAFQDTVRKEMT 199
DB 920 RLMKKVFTDDVIKEMT 935

RESULT 29
SYFA_STRCO STANDARD; PRT; 358 AA.
ID SYFA_STRCO
AC O88055;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase alpha chain (SC 6.1.1.20)
DE (Phenylalanyl-tRNA ligase alpha chain) (PheRS).
GN PHES OR SC01595 OR SC195.17C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
RN NCBI_TaxID=1902;
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
```

```
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
Warren T., Wistritz A., Woodward J., Barrall B.G., Parkhill J.,
Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RT Nature 417:141-147 (2002).
RL CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
diphosphate + L-phenylalanyl-tRNA(Phe).
CC -1- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -1- SIMILARITY: Belongs to alpha chain subfamily 1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch)
CC EMBL; AL939109; CNA20806.1; ALT_INIT.
DR HSSP; P27001; LPYS.
DR HAMAP; MF_00281; -.
DR InterPro; IPR004188; Phe tRNA_synth_N.
DR InterPro; IPR004529; PheS.
DR InterPro; IPR002319; tRNA-synth_2d.
DR InterPro; IPR006195; tRNA_ligase_II.
DR Pfam; PF02912; Phe-tRNA-synth_N; 1.
DR Pfam; PF01409; tRNA-synth_2d; 1.
DR TIGRfam; TIGR00468; pheS; 1.
DR PROSITE; PS00862; AA tRNA LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Magnesium; Complete proteome.
FT METAL 262 262 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 358 AA; 39119 MW; 49CF6E0000E469C CRC64;

Query Match 7.9%; Score 84; DB 1; Length 358;
Best Local Similarity 23.4%; Pred. No. 13;
Matches 33; Conservative 26; Mismatches 50; Indels 32; Gaps 5;

QY 27 EINKAIDDAIAEQSETIDPMKVPDHAQK-----FERHVGIV--DEKGLAMENI 76
DB 4 EIEIRMRDEALAAFAAADSLDQAKVAHTGGASPLANREIGALPPQAKAEAGKRVG 63
QY 77 EARGL-----KMKRQGDANYKGEIVKAHLIGVHDDIVSMEYDLYKGLDLHP 127
DB 64 MARGAVNKALAAQBELEGAERDARVLVEAV-----DVTLPHD-RVPAGARHP 110

RESULT 30
DEOB_STRCO STANDARD; PRT; 392 AA.
ID DEOB_STRCO
AC Q99X76;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phosphotransferase (EC 5.4.2.7) (Phosphoheoxyribomutase).
GN DEOB OR DRM OR SAV0139 OR SA0134 OR MW0113.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
```


Db 120 TLRCKLLAQVSLGSPDRVGLGQQQLVETAKALNKQVRLILDEPTASLITEQETS 179
QY 189 IFQDTR 195
Db 180 ILDIIR 186

RESULT 32
CH60_ENTGE STANDARD; PRT; 540 AA.
AC O66194;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin 1 (Protein Cpn60) (Fragment).
GN GROEL OR GROEL OR MOPA.
OS Enterobacter gergoviae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=61647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1234;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008139; BAA35211.1; -;
CC HSSP; P06139; IJON.
CC
CC InterPro; IPR001844; Chaprin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
FT NON TER 540 540
SQ SEQUENCE 540 AA; 56617 MW; B6DA54F0F5F84BBD CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 540;
Best Local Similarity 18.2%; Pred. No. 23;
Matches 43; Conservative 41; Mismatches 85; Indels 67; Gaps 8;

QY 13 VASADPHYDKITEINKAIDAAIEQSETDPMKVPDPAKDFERHVGIVDFKGEIA 72
Db 107 VAGNPNM-----DLKRGIDKAVAA--AVEELKALSVPCSDSKAIAQVGTISANSDET 157
QY 73 MENIARGLKQMKRGDANVGEIGVKAHLIIGVHDD---IVSMEDYLA----- 120
Db 158 VGKLTAEAMDKV-----GKEGVITVEDGTGLDLDVVEGMQDRGVLSPFYFNKP 208
QY 121 -----KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITWTSFEVR 160

Db 209 ETGAVELDSPFFILLADKKISNIRELLPVLVEAVAKAGKELLIIAEDVEGEALATL----- 262
QY 161 QPANNVNHGGI-----SILDPFGVLSVLTAFQDTRKEMTKVLAFAFKRELEK 212
Db 263 -----VNTMRGIVKVAVKAPGFG--DRRKAMLDIATLTAGTVISSEIGMELEK 311

RESULT 33
CH61_VIBVU STANDARD; PRT; 546 AA.
AC O9ALA9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 60 kDa chaperonin 1 (Protein Cpn60) (GroEL protein 1).
GN GROEL OR GROEL OR GROEL OR VV11260.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RA Wong H.-C., Lu K.-H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Khee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.B.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY017169; AAG48876.1; -;
CC EMBL; AB016801; AAC09716.1; -;
CC HSSP; P06139; IGRU.
CC
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaprin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
FT CONFLICT 134 134 L -> M (IN REF. 1).
FT CONFLICT 220 220 I -> F (IN REF. 1).
FT CONFLICT 256 256 G -> D (IN REF. 1).
FT CONFLICT 272 272 K -> I (IN REF. 1).
FT CONFLICT 277 277 K -> I (IN REF. 1).
FT CONFLICT 388 388 E -> K (IN REF. 1).
FT CONFLICT 474 474 G -> A (IN REF. 1).
FT CONFLICT 545 545 M -> GM (IN REF. 1).
SQ SEQUENCE 546 AA; 57515 MW; 92869B7E517B3D2A CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 546;
Best Local Similarity 25.0%; Pred. No. 23;
Matches 42; Conservative 19; Mismatches 70; Indels 37; Gaps 5;

QY 6 LIAAVAFVAVSA-----DPIHYDKITEINKAIDDAIAAEQSETIDPMKV 51
Db 372 LAGGAVIKVGAATEVEMKEKORVEDALHATPAAVEGIVAGGVALIRASKI-----V 427
QY 52 PDHADKFERHVGIVDFKGLAMRNIEARGLKQKMGROGDA-----NVKGEIGIVKAHLI 105
Db 428 DLQGDNEEQNGVI-----RVALRAMAPLRQITKNAGDESVVANNVRAGEGSGYNAAT 482
QY 106 GHVDIVSMVEYDLAYKGLDHPHTTHVISIDQFPVVALSLEISDEGNT 153
Db 483 GVVGDMLEM-----GILDPTKVTTSALQPAASVAGLMITTEAVMT 522

RESULT 34
CH60_YERPE STANDARD; PRT; 548 AA.
AC Q821Y3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 60 kDa chaperonin (protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR MOPA OR YPO0351 OR Y0609.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Farraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ141442; CAC89210.1; -;
CC EMBL; AE013663; AAM84197.1; -;
CC PIR; AG0043; AG0043.
CC HAMAP; MF 00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC InterPro; IPR008950; GroEL-ATPase.

DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 548 AA; 57431 MW; 6E33B1F889B79E3A CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 548;
Best Local Similarity 18.1%; Pred. No. 23;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VASADPIHYDKITEINKAIDDA-IAAIFQSETIDPMKVPDHDADKFERHVGIVDFKGL 71
Db 107 VAAGNPNM-----DLKRGIDKAVIAAVEE---LKKLSVPCSKAKAQAQVGTISANSDS 156
QY 72 AMRNIEARGLKQKMGROGDANVKGEGIVKGAHLHGVHDD---IVSMVEYDLAY----- 120
Db 157 TVGELIAQAMEKV-----GREGVITVEEGSLQDELVDVVGMGFDRGYLSPYFINK 207
QY 121 -----KLGDLHPTTHVISIDQFPVVALSLEISDEGNTIMTSFEV 159
Db 208 PETGSIELESPFILLADKKISNIREMLPVLEAVAKAGKPLLIITAEVDGEALATL----- 262
QY 160 RQFANVNVNHGGL-----SILDPFPGVLSDLVITAFQDTVRKMTKVLAPAFKRELEK 212
Db 263 -----VNTMRGIVKVAAPGFG---DRKAMQLQDIATLTAGTVISEIGLELEK 311

RESULT 35
CH60_YEREN STANDARD; PRT; 550 AA.
AC P492L9;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (protein Cpn60) (groEL protein) (Heat shock protein
DE 60) (Cross-reacting protein antigen).
GN GROEL OR GROEL OR MOPA OR HSP60 OR CRPA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y-108C / Serotype O:3;
RA Hafner C.E., Roggenkamp A.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA / Serotype O:8;
RA Autenrieth I.B., Noll A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype O:3;
RX MEDLINE=94248434; PubMed=8190995;
RA Yamamoto T., Miura H., Ohsumi K., Yamaguchi H., Taguchi H.,
RA Ogata S.;
RT "Cloning and nucleotide sequence analysis of immunodominant
RT heat-shock protein of Yersinia enterocolitica."
RL Res. Microbiol. 144:1691-701(1993).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ141442; CAC89210.1; -;
CC EMBL; AE013663; AAM84197.1; -;
CC PIR; AG0043; AG0043.
CC HAMAP; MF 00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC InterPro; IPR008950; GroEL-ATPase.

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

```

CC EMBL; X68526; CAA48539.1; -.
CC EMBL; X82212; CAA57694.1; -.
CC EMBL; D14078; BAA03164.1; -.
CC PIR; S26423; S26423.
CC HSSP; P06139; IJON.
CC HAMAP; MF 00600; -.
CC InterPro; IPR001844; Chapnain Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC InterPro; IPR008950; GCPEL-Antase.
CC Pfam; PF00118; Cpn60_TCF1.1
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chapterone; ATP-binding; Heat shock; Antigen.
FT VARIANT 2 355 355 E -> P (IN STRAIN WA).
FT VARIANT 16 17 ML -> IV (IN REF. 3).
FT CONFLICT 135 135 S -> Y (IN REF. 1).
FT CONFLICT 181 181 S -> P (IN REF. 1).
FT CONFLICT 297 297 A -> V (IN REF. 1).
FT CONFLICT 346 346 V -> L (IN REF. 3).
FT CONFLICT 347 347 A -> T (IN REF. 1).
FT CONFLICT 401 401 H -> D (IN REF. 3).
FT CONFLICT 423 423 A -> R (IN REF. 3).
FT CONFLICT 473 482 AGSSSYGNA -> RVQVATWS (IN REF. 3).
FT SEQUENCE 550 AA; 57527 MW; 226944B18E39DA4F CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 550;
Best Local Similarity 18.1%; Pred. No. 24;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVGADPHYDKITEINKAIDDA-IAAEQSEITDPMKVPDHDADKFERHVGIVDFKGL 71
DB 107 VAGNPNP-----DLKRGIDKAVIAAVEB---LKKLSVPCDSKAIAQVGTISANS 156
QY 72 AMRNIERAGLQKMQKQGDANVKYEGEIVKAHLHLLGVHDD---IVSMYDILAY----- 120
DB 157 TVGELLIAQMEKV-----GREGVITVEGSGLODELVDVVGQDFRGYLSPPYFINK 207
QY 121 -----KLGDLHPTTHVISDIQDFVVALSLEISDEGNTWTSFEV 159
DB 208 PETGSIELESPFILLADKKISNIREMLFVLEAVAKAGKPLLIITAEVDVEGALATL----- 262
QY 160 RQFANVNHIGGL---SILDPIFGVLSVLTALFQDVTVRKMTKVLAPAPKKELEK 212
DB 263 -----VNTMRGIVKVAAPGFG---DRKAMLDIATITAGTVISEIGILELEK 311

RESULT 36
RS4E METJA STANDARD; PRT; 244 AA.
AC P54039;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4e.
GN RPS4E OR MJ0468.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

```

```

RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).
CC -!- SIMILARITY: Belongs to the S4E family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67497; AAB98457.1; -.
CC PIR; D64358; D64358.
CC TIGR; MJ0468; -.
CC HAMAP; MF 00485; -.
CC InterPro; IPR005824; KOW.
CC InterPro; IPR006646; KOW_sub.
CC InterPro; IPR000876; Ribosomal_S4E.
CC InterPro; IPR002942; S4.
CC Pfam; PF00467; KOW; 1.
CC Pfam; PF00900; Ribosomal_S4e; 1.
CC Pfam; PF01479; S4; 1.
CC ProDom; PD002667; Ribosomal_S4E; 1.
CC SMART; SM00739; KOW; 1.
CC SMART; SM00363; S4; 1.
CC PROSITE; PS00528; RIBOSOMAL_S4E; 1.
CC PROSITE; PS00889; S4; 1.
KW Ribosomal protein; RNA-binding; Complete proteome.
FT DOMAIN 43 108 S4 RNA-BINDING.
FT SEQUENCE 244 AA; 27658 MW; 6759485019B86AC7 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 244;
Best Local Similarity 21.5%; Pred. No. 10;
Matches 42; Conservative 30; Mismatches 87; Indels 36; Gaps 6;

QY 6 LIAAFAVAVGADPHYDKITEINKAIDDAIAEQSEITDPMKVPDHDADKFERHVGIV 65
DB 45 LLLIVRDILKADNAREAKIKKGVLDVGRVKEKLPVGLMDVSLPDANENRVLP 104
QY 66 DPKGELAVRNIEARGLKQKQGDANVKYEGEIVKAHLHLLGVHDD-----DIVSMEY 116
DB 105 DRKGRIKLPTEPNPDVCLKIKKNTVIKG-----GHIQLNLDGRNIVIKVSDPTKAE 158
QY 117 DLAYKGLDHPHTHVISDIQDFVVALSLEISDEGNTWTSFEVQFANV--NHIGLSI 174
DB 159 DV-YKTGD-----TLISIEQEIKAHPEVKGKLVAYITGGKHVDFAK 201
QY 175 LDPI--FGVLSVLT 187
DB 202 IVEIERGGIYPIVT 216

RESULT 37
APEA_CLOAB
ID APEA_CLOAB STANDARD; PRT; 465 AA.
AC Q97K30;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable M18-family aminopeptidase 1 (EC 3.4.11.-).
GN APEA OR CAC1091.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]

```

```

SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D.J., Hitti J., Wolf Y.I.,
RA Tausov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -I- COFACTOR: Zinc (By similarity).
CC -I- SIMILARITY: Belongs to peptidase family M18.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL: AE007625; AAK79065.1; -.
DR PIR: F97034; F97034.
DR HAMAP: MF_00466; -.
DR InterPro: IPR001948; Peptidase_M18.
DR Pfam: PF02127; Peptidase_M18; 1.
DR PRINTS: PR00932; AMINO1PTASE.
DR Hypothetical protein; Hydrolase; Aminopeptidase; Metalloprotease;
KW Zinc; Complete proteome.
KW METAL 105 105 ZINC (POTENTIAL) .
FT METAL 180 180 ZINC (POTENTIAL) .
FT METAL 441 441 ZINC (POTENTIAL) .
SQ SEQUENCE 465 AA; 52011 MW; CB4C6D0A54C0A439 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 465;
Best Local Similarity 28.4%; Pred. No. 21; Gaps 10;
Matches 4; Conservative 27; Mismatches 49; Indels 30; Gaps 10;

QY 74 RNIE---ARGLKQMKRGD---ANVK-----GEGIVKAAHLITGVHDDI--VSMEY 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58 RNIEDILAKG-EFLKEGDKVYANNRGKGLIMFLGKEPLYTGFKILGAHIDSPRLDLKQ 115

QY 117 DLAYKLGDLHP-TTHVLSIDTQDFV-VALSLEISDEGNITWTSFVROFANVNVHIGLSI 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 NPLYETDLMLETHYGGYGGIKYQWVTLPLAI-HGVI-----VKDGTIVNVCVGEDD 167

QY 175 LDPIFGVLSDLTAIFODTVRKEMTKVL 202
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 168 NDFVFGV-SDILVHLASEQLEKASKVI 194

RESULT 38
PUR9 SYNEL STANDARD; PRT; 518 AA.
AC Q8DI5;
CT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bifunctional purine biosynthesis protein purH [Includes:
DE Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (AICAR transformylase); IMP cyclohydrolase (EC 3.5.4.10) (inosinase)
DE (IMP synthetase) (ATIC)].
GN PURH OR TLR1547
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22235144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Kawanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,

```

Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
"Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.",
DNA Res. 9:123-130(2002).
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-
phospho-D-ribose)imidazole-4-carboxamide = tetrahydrofolate + 5-
formamido-1-(5-phospho-D-ribose)imidazole-4-carboxamide.
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-
phosphoribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: De novo purine biosynthesis; ninth step.
CC -1- PATHWAY: De novo purine biosynthesis; tenth step.
CC -1- DOMAIN: The IMP cyclohydrolase activity resides in the N-terminal
region (by similarity).
CC -1- SIMILARITY: Belongs to the purH family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; AP005374; BAC0909.1; -.
CC DR HAMAP; MF 00139; -. 1.
CC DR InterPro; IPR002695; AICARFT_IMPChas.
CC DR InterPro; IPR004362; MGS_like.
CC DR Pfam; PF01808; AICARFT_IMPChas; 1.
CC DR Pfam; PF02142; MGS; 1.
CC DR ProDom; PD004666; AICARFT_IMPChas; 1.
CC DR TIGRFAMs; TIGR00355; purH; 1.
CC KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;
CC KW Complete proteome.
CC SQ SEQUENCE 518 AA; 55271 MW; F1010495BDCFE6 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 518;
Best Local Similarity 23.5%; Pred.No. 24;
Matches 52; Conservative 34; Mismatches 89; Indels 46; Gaps 10;

QY 11 AFVAVSADRIHYDKITEINKATDDAIAIEQSETIDPMKVPDADKFEHVGIVDFKGE 70
DB 140 AHVTLVDSQSYTYLQEL-RLYGDQPAPFLACAAQAFALTAHYDQ-----AIAEVLQK 193
QY 71 LAMRNTEARGL-----KMKRQGDA-----NVKG--EEGIVKAHLIG---VH 108
DB 194 VTAAGTEPELTPVPFHLTGKQKQVLYAGENPHQRASWVCGAHPRGWAAAHLLQKLSY 253
QY 109 DDIVSMEYD---LAYKLGDLHPHTTWSIDQDFVALSLEISDEGNITMTSPFVROFANV 165
DB 254 NLLDLEAARGVISEFLGSDSPAIVIKHTNPGVA-----EGKTLVEAYERAFADR 306
QY 166 VNHIGLSILD-PIFGLVDLVTAFQDVTVRKEMTKVLAPA 205
DB 307 VSAFGGIVALNRPLDGATAEALTRTFLEC-----VVAFA 340

RESULT 39
CH60_BACFO STANDARD; PRT; 543 AA.
ID CH60_BACFO
AC P81284; Q9X6Y3;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaparonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR MOPA.
OS Bacteroides forsythus.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Tannerella.
OX NCBI_TaxID=28112;
RN [1]
RP
RC SEQUENCE FROM N.A.
RC STRAIN=ATCC 43037;
RX MEDLINE=99452336; PubMed=10524765;


```

RA Reid H.I., Riggio M.P.;
RT "Identification and nucleotide sequence of the heat shock protein 60
RL (GroEL) gene of Bacteroides forsythus.",
RN DNASeq. 9:359-364 (1998).
RP [2]
RC SEQUENCE OF 1-39.
RD STRAIN=ATCC 43037;
RA Sojar H.T., Glurich I.B., Genco R.J.;
RL Submitted (APR-1998) to Swiss-Prot.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ006516; CAB43992.1; --
DR HSSP; P06139; IGEL; --
DR HAMAP; MF_00600; --; 1.
DR InterPro; IPR001844; Chapnrm Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN60_CPN60; 1.
DR Chaperone; ATP-binding.
FT INIT MET 0
FT CONFLICT 3 23 EIKFDNARDLLKGVDELAN -> DVKFGDARVRLGRV
FT NVALD (IN REF. 2).
FT ILE -> VLD (IN REF. 2).
FT SEQUENCE 543 AA; 57840 MW; 3AD767AF3494951 CRC64;
Query Match 7.8%; Score 83; DB 1; Length 543;
Best Local Similarity 22.3%; Pred. No. 26;
Matches 47; Conservative 36; Mismatches 72; Indels 56; Gaps 11;
QY 13 VAVSADPHYDKITEEINKAIDDAIAIEOSEITIDPMKVPDHDKPERHVGIVDFKGELA 72
Db 105 VTGANPW-----DKRGIDKAVKVKVESIASQSEAVGTNMDRE-HVAKISANGDEG 156
QY 73 MRNTEARGLKMKRQGDANV---KGE-----GIVKAHLILGVHDDIVSME- 115
Db 157 IGKLTAEAMQVKKEGIVTVEEAKGTETTVVEVVEGQFDRGYISAFV-----TDTSEKMET 212
QY 116 -----YDLAYK-----LGLDHPHTTHVTSIDIQDFVALSLSEISDEGNITMTSFEVRQFANV 165
Db 213 QFENPYILYDKTSVLKDLPLILEQVQSGRALLIATIEDIDSEALTL-----V 262
QY 166 VNHI--GGI---SILDPIFGVLSVDLTAIFQD 192
Db 263 VNRLRGGLKVCVAKPAGFG---DRRKAMLED 290
RESULT 40
DNAA AGRTS
ID DNAA AGRTS STANDARD; PRT; 633 AA.
AC P50019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
GN DNAA OR ATU0122 OR AGR_C_195.

```

```

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
NCBI_TaxID=1176299;
RP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96011387; PubMed=7592349;
RA Segal G., Ron E.Z.;
RT "The dnaK operon of Agrobacterium tumefaciens: transcriptional
RT analysis and evidence for a new heat shock promoter.",
RL J. Bacteriol. 177:5952-5958 (1995).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Roux G., Saepinmichak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.",
RL Science 294:2317-2323 (2001).
RP [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollman C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.",
RL Science 294:2323-2328 (2001).
CC -!- FUNCTION: Acts as a chaperone (By similarity).
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87113; CAA60592.1; --
DR EMBL; AE008986; AAL41147.1; --
DR EMBL; AE007953; AAK85942.1; --
DR PIR; AB2591; AE2591.
DR PIR; E97373; E97373.
DR PIR; I39585; I39585.
DR HSSP; P04475; IDG4.
DR HAMAP; MF_00332; --; 1.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS10036; HSP70_3; 1.
DR Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MCD RES 198
FT CONFLICT 293 293 L -> P (IN REF. 1).
FT CONFLICT 345 345 Q -> R (IN REF. 1).
FT CONFLICT 463 463 G -> A (IN REF. 1).
FT SEQUENCE 633 AA; 68080 MW; D53CC7E6EB5B5083 CRC64;

```

[illegible]

```

RESULT 42
CBIO_METJJA STANDARD; PRT; 279 AA.
ID CBIO_METJJA Q58488;
AC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable cobalt transport ATP-binding protein cbio.
DE CBIO OR MJ1088.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
[1]
SEQUENCE FROM N.A.
RN STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: May be involved with cobalt transport in association
CC with cobalamin biosynthesis (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/ebis/sib.ch/
CC or send an email to license@sib.ch).

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67551; AAB99089.1; --
 DR PIR; G64435; G64435.
 DR HSP; Q58663; IG6H.
 DR TIGR; MJ1089; --
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005876; CblN_ATP.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMS; TIGR01166; cbio; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Cobalamin biosynthesis; Cobalt transport; Transport; ATP-binding;
 KW Complete proteome.
 NP BIND 37 43 ATP (POTENTIAL).
 FT SEQUENCE 279 AA; 31120 MW; 7A2DF31BE68CFA81 CRC64;
 SQ

Query Match	7.7%;	Score 82.5;	DB 1;	Length 539;
Best Local Similarity	18.6%;	Pred. No. 28;		
Matches	44;	Conservative 43;	Mismatches 81;	Indels 69;
Gaps	9;			
QY	13	VAYSADPHYDKITEINKAIDDA-IAALEQSEETIDPMKVPHADKPERHVGIVDFKGL	71	
Db	107	VAGWNP-----DLKRGIDRAVIAVEELKTL--LSVCSDSKAIQAQVGTISANSDE	156	
QY	72	AMRNTBARGLKMQRQGDANVKGEGIVKAHLILGHDD--IVSMEYDLAY-----	120	
Db	157	TVGKLIAEAMDKV-----GKGQVITVEDGTGLELDELVDVVGQMFQRGYLSVPYINK	207	
QY	121	-----KLGDILPHTHVISDIQDFWALSLSIDEGNITMTSFEV	159	
Db	208	PETGAVELESPFILLADKKISIMREMLPYLEAVAKAGKPLLIAEVSVEGEAKATL-----	262	
QY	160	RFQANVWNHIGGL-----SILDPFGVLSVDLTAIFQDTPVRKEMTKVLAPAFKRELEK	212	
Db	263	-----VWNTMRGIVKVAAPKAPFG-----DERRAKMLQDIATLFGGTVISPEIGMELEK	311	

```

RESULT 44
CH60_KLEPL STANDARD; PRT; 540 AA.
ID CH60_KLEPL
AC O66212;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein) (Fragment).
OS GROEL OR MOF4 OR GROEL.
OC Klebsiella planticola (Raoultella planticola).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Raoultella.
OX NCBI_TaxID=575;
RN [1]
RP SEQUENCE FROM N.A.
RR STRAIN=JCM 7251;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB0081148; BAA25229.1; -.
CC DR HMAP; P06139; IJON.
CC DR HMAP; MF 00600; -.
CC DR InterPro; IPR001844; Chaperin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCF-1.
CC DR InterPro; IPR008950; GroEL-ATPase.
CC DR Pfam; PF00118; cpn60_TCF1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCP1.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC DR Chaperone; ATP-binding.
CC DR NON_TER 540
CC DR SEQUENCE 540 AA; 56655 MW; 9E118C11ABC2430A CRC64;
CC FT
Query Match 7.7%; Score 82.5; DB 1; Length 540;

```

Query Match	7.7%;	Score 82.5;	DB 1;	Length 279;
Best Local Similarity	26.9%;	Pred. No. 13;		
Matches 49;	Conservative 29;	Mismatches 69;	Indels 35;	Caps 10;
QY	15	VSADPIHYDKITTEINKAIDDAIAAI--EQSETIDPMKVPDHADKFERHVVYDFKGEA	72	
Db	102	VAFGPLNLGLPKEVEKRVKEALKAVGMGEFENKPPHLLSGGOKKRVATAGILAMQPEVI	161	
QY	73	MRIIEARGH-----KMKRQGDANVKGEEI VKAHLILGVHD-DIVSMYEDLAYKLGD-	124	
Db	162	VLDEPTAGLDPPVCASKIMKLLYLNKKG-----WTIIISTHDVLDVPVYADKVVYMDG	215	
QY	125	--LHPTT--HVISDIODF-----VVALSLBI-SDEGNITM-----TSFEVROFANVN	167	
Db	216	KILKGTPTKEVSDVETIRKANLPRVAHLIEILNKNONIPIEWGFTICEVRR--NIVN	273	
QY	168	HI	169	
Db	274	YL	275	

```

RESULT 43
CH60_ENTAM
ID CH60_ENTAM STANDARD; PRT; 539 AA.
AC O66136;
DT DT 30-MAY-2000 (Rel. 39, Created)
DT DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment) .
DE GROEL OR GROEL OR MOFA.
OS Enterobacter amnigenus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
NCBI_TaxID=61646;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=JCM 1237;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species." ;
RL J. Gen. Appl. Microbiol. 43:355-361(1997) .
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB008140; BAA25213.1; -.
CC HSPSP; P06139; IJON.
CC HAMAP; MF_00600; -. 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCF-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60_TCF1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC NON TER 539 539
CC SEQUENCE 539 AA; 56657 MW; 09E701CB2B1DBA17 CRC64;
SQ

```

	Matches	43; Conservative	44; Mismatches	81; Indels	69; Gaps	9;
Qy	13	VAVSADPIHYDKTEINKAIDDA-TAAETQSTIDPMKUPDHADKPEERHUGIVDFKGEL	71			
Db	107	VAAGMNP-----DLKRGIDKAAVAAVEE-----LKKLSVPCSDSKAIAQVGTISANDE	156			
Qy	72	AMENIEARGLKQMKROGDANVKGEEGIVKAHLIIGVHDD---IVSMYDIAY-----	120			
Db	157	TVGKLIAEAMEKV-----GKEGIVTVBEETGLQDELDDVVEGMOFDRGLSPYFINK	207			
Qy	121	-----KLGDLPHTTHVLSIDIQDFVVALSLEISDEGNIIMTTSFEV	159			
Db	208	PETGSVELSPFTLLADKKLSNIRELLPVLEAVAKAGKPLIIVAEDVEGALATL-----	262			
Qy	160	RQFANVNVHIGGL-----SILDPIFGVLSVDLVLTAIQDITVRKEMTKVLAPAKRELEK	212			
Db	263	-----VVNTRRGIVKVAANVAKPFG---DRRKAMLQDIATLTAGTIVSEIGLELEK	311			

```

CC60_XANAC STANDARD; PRT; 546 AA.
ID AC Q8PP21;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL OR XAC0542.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=306 / ATCC 13902 / XV 101;
RC MEDLINE=22022145; PubMed=12024217;
RX da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bartolini M.C., Canargo L.E.A., L.P.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergro F., Clispina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggig C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locai E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Mencia C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sb-sib.ch).
CC -----
CC ENBL; AE011680; AM35431.1; --
DR HAMAP; MF 00600; -- 1.
DR InterPro; IPR001844; Chpnrln Cpn60.

```

DR HAMAP; MF_00600; -; 1.
DR InterPro; IPR001844; Chaprinin_Cpn60.

```
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXCP1.
DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 546 AA; 57166 MW; 07D167B4347D5C9 CRC64;

Query Match 7.7%; Score 82.5; DB 1; Length 546;
Best Local Similarity 20.8%; Pred. No. 28;
Matches 50; Conservative 40; Mismatches 79; Indels 71; Gaps 11;

QY 13 VAVSADPIHYDKTEENKAIIDAI-AAIEQSETIDPMKYPDHRADKPERHVGIVDFKGL 71
D 107 VAAGNFM-----DLRGIDQAVKAAVVELKNISK---PTTDDKAIAQVGTISANSD 156
QY 72 AMRNIARGLKMQRQGDANVKGEGIVKAHLIGVH---DDIVSMYDILAY----- 120
D 157 SIGNIIAEMKVV-----GREGVITVEEGSGLENELDVVEGMQDFRGYLSPIFINN 207
QY 121 -----KLGDLHPTTH--VISDIQDFVAL-----SLEISDEGNTITMSTFEV 159
D 208 QQSQSADLDDPFILLHDKISNVRLDLPVLEGVAKAGKPLLIIVAEVEGEALATL----- 262
QY 160 QFANVNHIGGL-----SILDPIFG-----VLSDLTAIFQDTVRKEMTKVLAFAKREL 210
D 263 -----VNTINGIKVAVKAPGFDRRKAKLEDMAVLGTGTIVSEVGLAEKATIKOL 317

RESULT 47
COA2 HUMAN STANDARD; PRT; 2493 AA.
AC 000763; Q16852;
DT 01-NOV-1997 (Rel. 35, Created)
DI 15-MAR-2004 (Rel. 43, Last sequence update)
DE Acetyl-CoA carboxylase 2 [EC 6.4.1.2] (ACC-beta) [Includes: Biotin
GN carboxylase (EC 6.3.4.14)].
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
[1] SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=97256787; PubMed=9099716;
RA Abu-Elsheiga L., Almaraz-Ortega D.B., Baldini A., Wakil S.J.;
RT "Human acetyl-CoA carboxylase 2. Molecular cloning, characterization,
RL J. Biol. Chem. 272:10669-10677 (1997).
[2] SEQUENCE OF 1349-2134 FROM N.A.
RC TISSUE=Adipose tissue;
RX MEDLINE=9626506; PubMed=8670171;
RA Widmer J., Fasshi K.S., Schlichter S.C., Wheeler K.S., Crute B.E.,
RA King N., Nulle-Mcmenemy N., Noll W.W., Daniel S., Ha J., Kim K.-H.,
RA Witters L.A.;
RT "Identification of a second human acetyl-CoA carboxylase gene.";
RL Biochem. J. 316:915-922 (1996).
CC -!- FUNCTION: ACC-beta may be involved in the provision of malonyl-CoA
CC or in the regulation of fatty acid oxidation, rather than fatty
CC acid biosynthesis. This protein carries three functions: biotin
CC carboxyl carrier protein, biotin carboxylase, and
CC carboxyltransferase.
CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
CC + malonyl-CoA.
CC -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)
CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.
CC -!- COFACTOR: Biotin.
CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first (rate-limiting)
CC step.
```

```

QY 143 S-----LEISDEGNIWTSFEVROFANVNVHIGLSIL 175
DB 2349 TFLYWLRLLELLEDO-----VKOEILQAGSELGHVHIOHML 2384

RESULT 48
DLDH CHLPN STANDARD; PRT; 461 AA.
AC Q9Z773; Q9Q0E6;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dihydrolipoamide dehydrogenase (EC 1.8.1.4) (E3 component of 2-
DE oxoglutarate dehydrogenase complex).
GN LPDA OR CPN0833 OR CPl037 OR CPB0862.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RC MEDLINE=99206606; PubMed=10192388;
RX Kaiman S., Mitchell W., Marathe R., Lammel C., Pan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RN [2]
RN Nat. Genet. 21:385-389(1999).
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher M., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RN [3]
RN Nucleic Acids Res. 28:1397-1406(2000).
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RN [4]
RN Nucleic Acids Res. 28:2311-2314(2000).
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RA Shirai M.;
RT "Genomic sequence comparison of two unrelated isolates of Chlamydia
RT pneumoniae from Japan and U.S.";
RN [5]
RN Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN-TW-183;
RX Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis.";
RN [6]
RN Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:
CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3) (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Dihydrolipoamide + NAD(+) = lipoamide + NADH.
CC -1- COFACTOR: Binds 1 FAD per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- MISCELLANEOUS: The active site is a redox-active disulfide bond.

```

```

CC -1- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide
CC oxidoreductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB001653; AAD18970.1; -
CC EMBL; AB002261; AAF38812.1; -
CC EMBL; AP002548; BAA99041.1; -
CC EMBL; AB035943; BAA88651.1; -
CC EMBL; AB017160; AAP98791.1; -
CC PIR; C72031; C72031.
CC PIR; G86594; G86594.
CC HSSP; F14218; ILPPF.
CC PHCI-2DPAGE; Q92773; -
CC TIGR; CP1037; -
CC InterPro; IPR001327; FAD_pyr_redox.
CC InterPro; IPR000815; Hg_reductase.
CC InterPro; IPR006258; Lipoamide_dh.
CC InterPro; IPR001100; Pyr_redox_dim.
CC InterPro; IPR004099; Pyr_redox_dim_2.
CC InterPro; IPR000103; Pyridine_redox_1.
CC Pfam; PF00070; Pyr_redox; 1.
CC Pfam; PF02852; Pyr_redox_dim; 1.
CC PRINTS; PR00368; FADPNR.
CC PRINTS; PR00945; HGRDASE.
CC PRINTS; PR00411; PNDRDTASEI.
CC PRINTS; PR00469; PNDRDTASEII.
CC ProDom; PD000139; FAD_pyr_redox; 1.
CC TIGRFAMs; TIGR01350; lipoamide_dh; 1.
CC PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
CC Redox-active center; Glycolysis; Oxidoreductase; NAD; Flavoprotein;
KW FAD; Complete proteome.
FT NP_BIND 34 42 FAD (ADP PART) (BY SIMILARITY).
FT DISULFID 42 47 REDOX-ACTIVE (BY SIMILARITY).
FT ACT_SITE 436 436 K -> T (IN REF. 5).
FT CONFLICT 64 64
FT SEQUENCE 461 AA; 49166 MW; CADD4483E758FFAD CRC64;
SQ
Query Match 7.7%; Score 82; DB 1; Length 461;
Best Local Similarity 20.4%; Pred. No. 25;
Matches 46; Conservative 44; Mismatches 59; Indels 66; Gaps 11;
QY 23 DKITEE-----INKAIDDAIAAEQSEIDPMKVPDHADKFE-----RH 61
DB 219 NKFTKQIGIRILTKA---SISAIEESQNOVRITVNDQVEEFDYLVLAIGRQFNATASIGLDN 275
QY 62 VGIV-DFKGLAMRNIEARGLKQKRGD-----ANVKGEGIVKAHLILGVHDDIVS 113
DB 276 AGVIRDRGVIPVDVETWRTNVPNIYAIGDITGKLLAHVASHQGVIAKNISGHE---V 332
QY 114 MEYDLAYKLGDLHPHTTHVISDIQDFVVALSLEISDEGNI---TMTSFEVRQ----- 161
DB 333 MDYSAIPSVITHP-----EIAMVGLSLQEAQQNLPAKLTFFPKAIGKAVLGAAS 384
QY 162 --FANVNH-----IGGSILIDPFVGLSDVLTAFQDTVRKEMT 199
DB 385 DGFAAIVSGHEITQOILGAYVIGPHASSLIGEMTL-----AIRNELT 425
RESULT 49
MYS3_SCHPO STANDARD; PRT; 2104 AA.
ID MYS3_SCHPO STANDARD;
AC O14157; O42730; Q9UU49;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin type II heavy chain 2.

```

```

GN MYO3 OR MYP2 OR SPAC4A8.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98119390; PubMed=9459302;
RA Motegi F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
RT "Identification of Myo3, a second type-II myosin heavy chain in the
RL fission yeast Schizosaccharomyces pombe.";
RN FEBS Lett. 420:161-166(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98062346; PubMed=9398685;
RA Bezanilla M., Forsburg S.L., Pollard T.D.;
RT "Identification of a second, myosin-II in Schizosaccharomyces pombe:
RL Myo3p is conditionally required for cytokinesis.";
RN Mol. Biol. Cell 8:2693-2705(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21849401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds R., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Beck A., Lehner H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jinenz J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RN Nature 415:871-880(2002).
RN [4]
RP SEQUENCE OF 1871-2078 FROM N.A.
RX STRAIN=968 b50;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RL fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RN Genes Cells 5:169-190(2000).
CC -!- FUNCTION: Stabilizes the F-actin cables forming the F-actin ring
CC that surrounds the nucleus during interphase. May work in
CC conjunction with myo2.
CC -!- SUBUNIT: Binds to cdc4 and rlc1.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AB007633; BAA24579.1; -
DR EMBL; AF029788; AAC04615.1; -
DR EMBL; Z88782; CAB11475.1; -
DR EMBL; AB027812; BAA87116.1; -
DR FIR; T38774; T38774.
DR HSP; P08799; 1MND.
DR GenesDB SPombe; SPAC4A8.05c; -.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; Myosin_N.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
KW Myosin; Coiled coil; ATP-binding; Actin-binding.
FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
FT NP_BIND 646 660 ATP (POTENTIAL).
FT CONFLICT 182 189 D -> G (IN REF. 2).
FT CONFLICT 1193 1193 E -> G (IN REF. 2).
FT CONFLICT 1304 1304 E -> K (IN REF. 2).
FT CONFLICT 1344 1344 G -> D (IN REF. 2).
FT CONFLICT 1420 1420 G -> D (IN REF. 2).
SQ SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;
Query Match 7.7%; Score 82; DB 1; Length 2104;
Best Local Similarity 24.8%; Pred. No. 1.5e+02;
Matches 36; Conservative 25; Mismatches 54; Indels 30; Gaps 6;
QY 24 KITEINKAIDDAIAAIQSTIDPMKYPDHADKFERHGVGVDFKGLAMEN----- 75
DB 1864 EVSEQLNKKFDSSARPP-----DEMEKEVITLTKSNIAQKDDLLSLVE 1908
QY 76 -IEARGLKQMRQGDVANKYEGEVKVAHLIGVHDDIVMEYDIAKGLDLPHTHWISD 134
DB 1909 RIKQIEMFALTKQDSNNHRENL-QLHRLGV---LQKEKKDLKLFLLDLTKTYPST 1964
QY 135 IQDFVVALSLISD-EGNITWTSFE 158
DB 1965 SKD-VRMLOKQISDLEASFAASDIE 1988
RESULT 50
CH60 PANAN
ID CH60 PANAN STANDARD; PRT; 541 AA.
AC O66218;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOPA.
OS Pantoea ananas [Erwinia uredovora].
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 6986;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RL related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RN J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB008151; BAA29235.1; --
 DR HSSP; P06139; 1DK7.
 DR HAMAP; MF_00600; --; 1.
 DR InterPro; IPR001844; Chaprinin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60_TCPI; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCPI.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 DR Chaperone; ATP-binding.
 DR NON_TER 541 541
 SQ SEQUENCE 541 AA; 56817 MW; 7111AA03A703E543 CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 541;
 Best Local Similarity 18.1%; Pred. No. 34;
 Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSEITDPMKVPDHPADKFERHVGIVDFKGL 71

Db 107 VAGGNPM-----DLKRGIDQAVAAVEE---LKLSPGSDSKAIAQVGTISANSD 156

QY 72 AMENIEARGLKQMKQGDANVKEGIVKAHLIGVHD---IVSNEYDLAY----- 120

Db 157 TVGTLIAQAMEKV-----GKEGVITVEGTGLQDELVDVVEGQFDRGLSPYFINK 207

QY 121 -----KLGDLHPHTHVISDIQDFVVALSLSEISDEGNITWTSPEV 159

Db 208 PETGAVELETPPILLADKKISNIREMLPVLEAVAKAGKPLLIADVEGEALATL----- 262

QY 160 ROPANYVNHIGL-----SILDPIFGVLSDLTAIFQDITVRKEMTKVLAFAFKRELEK 212

Db 263 -----VNTMTRGIVKVAAPGFG---DRRKAMLDIAILTGTGTVISEIGMELEK 311

RESULT 51

CH60_HAEDU STANDARD; PRT; 547 AA.

AC P31294;

DT 01-JUL-1993 (Rel. 26, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).

GN GROEL OR GROEL OR MOPA OR HD1784.

OS Haemophilus ducreyi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=730;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93014108; PubMed=1356926;

RA Parsons L.M., Waring A.L., Shayegani M.;

RT "Molecular analysis of the Haemophilus ducreyi groE heat shock

operon".

RL Infect. Immun. 60:4111-4118(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=3500HP / ATCC 700724;

RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,

RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;

RT "The complete genome sequence of Haemophilus ducreyi".

RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M91030; AAA24961.1; --
 DR EMBL; AE017155; AAB96536.1; --
 DR PIR; B49203; B49203.
 DR HSSP; P06139; 1GRL.
 DR HAMAP; MF_00600; --; 1.
 DR InterPro; IPR001844; Chaprinin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60_TCPI; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCPI.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 DR Chaperone; ATP-binding; Complete proteome.
 DR CONFLICT 37 N -> K (IN REF. 1).
 DR CONFLICT 37 A -> V (IN REF. 2).
 FT CONFLICT 509 509
 SQ SEQUENCE 547 AA; 57798 MW; B9DF93AD998B5C8 CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 547;
 Best Local Similarity 18.2%; Pred. No. 34;
 Matches 43; Conservative 44; Mismatches 82; Indels 67; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAIEQSEITDPMKVPDHPADKFERHVGIVDFKGLA 72

Db 107 VAGGNPM-----DLKRGIDKAAVAV--VEELKAISKCEKSEIEQVGTISANSD 157

QY 73 MNIEARGLKQMKQGDANVKEGIVKAHLIGVHD---DIV-SNEYDLAY----- 120

Db 158 VGKLIQAAMEKV-----GKEGVITVEDGTGLDLDVVEGQFDRGLSPYFINKP 208

QY 121 -----KLGDLHPHTHVISDIQDFVVALSLSEISDEGNITWTSPEVR 160

Db 209 EAGTVELDNPIIIVDKKISNIREILPVLEAVAKAGKPLLIADVEGEALATL----- 262

QY 161 QFANYVNHIGL-----SILDPIFGVLSDLTAIFQDITVRKEMTKVLAFAFKRELEK 212

Db 263 ----VNTMTRGIVKVAAPGFG---DRRKAMLDIAILTGTGTVISEIGMELEK 311

RESULT 52

CH61_VIEHA STANDARD; PRT; 548 AA.

AC Q83WI8;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE 60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1).

GN GROEL OR GROEL1.

OS Vibrio Harveyi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=669;

RN [1]

RP SEQUENCE FROM N.A.

RA Kuchanny-Ardigo D., Lipinska B.;

RT "Cloning and characterization of the groE heat shock operon of the

marine bacterium Vibrio Harveyi".

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Prevents misfolding and promotes the refolding and

proper assembly of unfolded polypeptides generated under stress

conditions (By similarity).

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

7 subunits (By similarity).

RESULT 55
GID_AQUAE
ID_ GID
AC 0669

```

RESULT 56
UNR_HUMAN
ID UNR_HUMAN STANDARD; PRT; 798 AA.
AC O75534; O94961; O9Y294;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UNR protein.
GN UNR OR D15155E OR KIA0885.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Pancreatic;
RA Song H., Peng Y., Zhou J., Huang Q., Dai M., Mao Y., Yu Y., Xu X.,
RA Luo M., Hu R., Chen J.;
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Cervix;
EX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altshul S.F., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Hopkins R.F., Jordan K.H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 1-498 FROM N.A. (ISOFORM SHORT).
RC TISSUE=Brain;
RA Xu W., Gibbs R.A.;
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP INTERACTION WITH UNRIP.
RX MEDLINE=99158740; PubMed=10049359;
RA Hunt S.L., Huan J.J., Totty N., Jackson R.J.;
RT "unr, a cellular cytoplasmic RNA-binding protein with five cold-shock
RT domains, is required for internal initiation of translation of human
RT rhinovirus RNA.";
RL Genes Dev. 13:437-448(1999).
CC -1- FUNCTION: RNA-binding protein. Required for internal initiation of
CC translation of human rhinovirus RNA.
CC -1- SUBUNIT: Interacts with UNRIP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS:

```

```

CC CC Event=Alternative splicing; Named isoforms=2;
CC CC Name=Long;
CC CC IsoId=O75534-1; Sequence=Displayed;
CC CC Name=Short;
CC CC IsoId=O75534-2; Sequence=VSP_001138;
CC CC Note=No experimental confirmation available;
CC CC -1- SIMILARITY: Belongs to the cold-shock domain (CSD) family.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AB020692; BAA74908.1; -
CC CC EMBL; AF077054; RAD2787.1; -
CC CC EMBL; BC032446; RAH32446.1; -
CC CC EMBL; AF070542; AAC28634.1; -
CC CC FIR; S29815; S29815.
CC CC HSSP; P15277; LMJC.
CC CC MIM; 191510; -
CC CC GO; GO:0008584; P: male gonad development; TAS.
CC CC InterPro; IPR002059; Cold_shock.
CC CC InterPro; IPR008994; Nucleic_acid_OB.
CC CC Pfam; PF00313; CSD; 7.
CC CC ProDom; PD000621; Cold_shock; 1.
CC CC SMART; SM00357; CSP; 5.
CC CC PROSITE; PS00352; COLD_SHOCK; 4.
CC CC RNA-binding; Repeat; Alternative splicing.
CC CC FT DOMAIN 26 87
CC CC CSD 1.
CC CC CSD 2. (INCOMPLETE).
CC CC CSD 3. (INCOMPLETE).
CC CC CSD 4. (INCOMPLETE).
CC CC CSD 5.
CC CC CSD 6.
CC CC CSD 7.
CC CC CSD 8.
CC CC CSD 9.
CC CC Missing (in isoform Short).
CC CC /ftid=VSP_001138.
CC CC E -> G (IN REF. 2).
CC CC QEIL -> TRNP (IN REF. 2).
CC CC T -> P (IN REF. 2).
CC CC IKVDVDRVMPFHFSEILDGNOL -> HPSVWIRECSVCSF
CC CC PLPVKFWMTSS (IN REF. 2).
CC CC SEQUENCE 798 AA; 88884 MW; 2ED0B32F33D454DA CRC64;
CC CC Query Match 7.6%; Score 81; DB 1; Length 798;
CC CC Best Local Similarity 22.9%; Pred. No. 58;
CC CC Matches 36; Conservative 29; Mismatches 56; Indels 36; Gaps 8;
QY 15 VSADPIHYDKITEINKAIDDAIAIEQSEITIDPKYDPDHAD-KFERHVGIV---DFKGE 70
Db 574 VSAEKVNK---THSVNGITEADPTIYSGKVRILRSVDPDTQTYQGMIEIVEGDMKGE 630
QY 71 LAMRNIEARGLKQKQGDANVKGEGIVKAKHLIGVHDDIVSMYDLA----- 119
Db 631 -----VYPFGIVGVKNGKDCIQGES--VKFQLCV-LQNAQTWAYNITPLRRATVECVK 682
QY 120 -----YKLGDLHPTTHVISDIQDFVVALSLISDE 149
Db 683 DQFGFNYGVGDSKKLPPHVKVQD---GIELQAGDE 716
RESULT 57
BFR_SERMA
ID_BFR_SERMA STANDARD; PRT; 159 AA.
AC O68935;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```
DE Bacterioferritin (BFR) (Cytochrome B-1) (Cytochrome B-557).
GN BFR.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RA Noorani S.M., Lindahl L., Zengel J.M.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May perform analogous functions in iron detoxification
CC and storage to that of animal ferritins (By similarity).
CC -!- COPACTOR: Binds 1 heme B (iron-protoporphyrin IX) group per dimer
CC (By similarity).
CC -!- SUBUNIT: Oligomer of 24 identical subunits (By similarity).
CC -!- MISCELLANEOUS: The di-iron binding site functions as active site
CC where iron ions are oxidized from iron(II) to iron(III) before
CC they are stored (By similarity).
CC -!- SIMILARITY: Belongs to the bacterioferritin family.
CC -!- SIMILARITY: Contains 1 ferritin-like diiron domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF058451; AAC14293.1; -.
DR HSPSP; P11056; IBCF.
DR InterPro; IPR002024; Bacterioferritin.
DR InterPro; IPR008331; Ferritin Dps.
DR InterPro; IPR009040; Ferritin_like.
DR Pfam; PF00210; ferritin; 1.
DR PRINTS; PR06001; BACFERRITIN.
DR PRODOM; PD002269; Bacterioferritin; 1.
DR TIGRFAMs; TIGR00754; bfr; 1.
DR PROSITE; PS00549; BACTERIOFERRITIN; 1.
DR PROSITE; PS00905; FERRITIN LIKE; 1.
KW Iron storage; Heme; Metal-binding; Iron.
FT DOMAIN 1 145 FERRITIN-LIKE DIIRON.
FT METAL 18 18
FT METAL 51 51 IRON 1 (BY SIMILARITY).
FT METAL 51 51 IRON 1 (BY SIMILARITY).
FT METAL 51 51 IRON 2 (BY SIMILARITY).
FT METAL 52 52 IRON 2 (BY SIMILARITY).
FT METAL 54 54 IRON 1 (BY SIMILARITY).
FT METAL 54 54 IRON 2 (BY SIMILARITY).
FT METAL 94 94 IRON 1 (BY SIMILARITY).
FT METAL 127 127 IRON 2 (BY SIMILARITY).
FT METAL 127 127 IRON 2 (BY SIMILARITY).
FT METAL 130 130 IRON 2 (BY SIMILARITY).
SQ SEQUENCE 159 AA; 18533 MW; 831A865EA98C382D CRC64;
Query Match 7.5%; Score 80.5; DB 1; Length 159;
Best Local Similarity 24.6%; Pred. No. 9.7; Mismatches 46; Indels 61; Gaps 8;
Matches 41; Conservative 19;
QY 24 KITSEINKAIDDAIAAEIQ-----SETIDPMKVPDHDADKPERHV 62
DB 6 KITAHKLKLNELVALNQYFLHARFKWGLRLNDKEYHESIDEMK---HADRYIERI 62
QY 63 ----GIYDFKGEIARNT-----EARGLKMKRGDANVKGEIGVKAHL 104
DB 63 LFLEGIPNLQ-DLCKLNGEDIEMLRSDLALELAGAKNL-REGIAYADSIHDYVSRDLM 120
QY 105 IGV-----HDDIVSMYDLYKGLDLPHTTHVISDIQDFVVALSLE 145
DB 121 IDILADEEHIDWLETDLIARLG-----IQNYAQAIL 156
RESULT 58
VATE SULTO
ID VATE_SULTO STANDARD; PRT; 194 AA.
AC Q971B8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V-type ATP synthase subunit E (EC 3.6.3.14) (V-type ATPase subunit E).
GN ATPB OR ST1435.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., JIN-NO K., TAKAHASHI M.,
RA SEKINE M., BABA S.-I., ANKAI A., KOSUGI H., HOSOYAMA A., FUKUI S.,
RA NAGAI Y., NISHIJIMA K., OTSUKA R., NAKAZAWA H., TAKAMURA M., KATO Y.,
RA YOSHIZAWA T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K.-I., MASUDA S., YANAGII M., NISHIMURA M., YAMAGISHI A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7."
RL DNA Res. 8:123-140(2001).
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SIMILARITY: Belongs to the V-ATPase E subunit family.
CC
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF000986; BAB66502.1; -.
DR HAMAP; MF_00311; -.
DR InterPro; IPR002842; ATPsynth_Esub.
DR Pfam; PF01991; vATP-synt_E; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport;
KW Complete proteome.
SQ SEQUENCE 194 AA; 22634 MW; 20F46AF95783449A CRC64;
Query Match 7.5%; Score 80.5; DB 1; Length 194;
Best Local Similarity 20.6%; Pred. No. 12;
Matches 41; Conservative 39; Mismatches 68; Indels 51; Gaps 9;
QY 22 YDKITEEINKAIDDAIAAEQSEITIDPMKVPDHDADKPERHVGVDFKGEIARNTAERGL 81
DB 26 FKILSEMNQIDEAYAEVYREYS---AKITDLVNKNNDRI-----RGEIARMEIENKRL 77
QY 82 KMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLYKGLDLPHTTHVISDIQDFVVA 141
DB 78 --ISKEMDYWENVKNAKSLYRFVKTD-----NYKKG-----LESI----- 113
QY 142 ISLEISDGNITMTSFVFRQFANVNH-----IGLSILDPIGVLSLSD----- 184
DB 114 ISREVSDEGSIIVCSPSDQKSIDIKKKKISKIIVWDEKIVGGIKYIPDKSLSKDFTLE 173
QY 185 -VLTAFODTVRKEMTKVL 202
DB 174 TILNQVFDI-IDKIAQIL 191
RESULT 59
CH60_ENTAG
ID CH60_ENTAG STANDARD; PRT; 539 AA.
AC O66200;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

```

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOPA.
OS Enterobacter agglomerans (Pantoea agglomerans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]_TaxID=549;
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1236;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008142; BAA25217.1; -
CC HSSP; P06139; 1JON.
CC DR HAMAP; MF 00600; -; 1.
CC DR InterPro; IPR001844; Chaprin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR InterPro; IPR008950; GroEL-ATPase.
CC DR Pfam; PF00118; cpn60_TCP1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCP1.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC FT NON_TER 539 539.
CC SQ SEQUENCE 539 AA; 56373 MW; 6583295F73299C3B CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 539;
Best Local Similarity 18.1%; Pred. No. 40;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPTHYDKITEINKAIDDAIA-AIQSEITIDPKVPDHDADKPERHVGIVDFKGL 71
DB 107 VAGNPNM-----DLKRGIDKAVTAABVE---LKALSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKQMRQGDANVKGEEGVKAHLIGVHDD---IVSMYEDLAY-----120
DB 157 TVGKLIAEAMDKV-----GREGVITVEDGTGLELDVVEGMDRGLSPYFINK 207
QY 121 -----KLGDLHPTHTVTSIDQIDFVVALSLEISDEGNTWTSFEV 159
DB 208 PETGAVELESFPILLADKKISNIREMLFVLEAVAKAGKPLVIAEDVEGEALATL-----262
QY 160 RQFANVNVHIGL-----SILDPIRGVLSDLVLTAFQDTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VVTMREGIVKVAAPKPGF---DRKAMLDQIATLTGTVISEEIGMELEK 311

RESULT 60
CH60 ENTAS
ID CH60 ENTAS STANDARD; PRT; 539 AA.
AC O66190;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).

GN GROEL OR GROEL OR MOPA.
OS Enterobacter agglomerans (Pantoea agglomerans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]_TaxID=549;
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1236;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008142; BAA25217.1; -
CC HSSP; P06139; 1JON.
CC DR HAMAP; MF 00600; -; 1.
CC DR InterPro; IPR001844; Chaprin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR InterPro; IPR008950; GroEL-ATPase.
CC DR Pfam; PF00118; cpn60_TCP1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCP1.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC FT NON_TER 539 539.
CC SQ SEQUENCE 539 AA; 56373 MW; 6583295F73299C3B CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 539;
Best Local Similarity 18.1%; Pred. No. 40;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPTHYDKITEINKAIDDAIA-AIQSEITIDPKVPDHDADKPERHVGIVDFKGL 71
DB 107 VAGNPNM-----DLKRGIDKAVTAABVE---LKALSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKQMRQGDANVKGEEGVKAHLIGVHDD---IVSMYEDLAY-----120
DB 157 TVGKLIAEAMDKV-----GREGVITVEDGTGLELDVVEGMDRGLSPYFINK 207
QY 121 -----KLGDLHPTHTVTSIDQIDFVVALSLEISDEGNTWTSFEV 159
DB 208 PETGAVELESFPILLADKKISNIREMLFVLEAVAKAGKPLVIAEDVEGEALATL-----262
QY 160 RQFANVNVHIGL-----SILDPIRGVLSDLVLTAFQDTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VVTMREGIVKVAAPKPGF---DRKAMLDQIATLTGTVISEEIGMELEK 311

RESULT 61
CH60 ENTAS
ID CH60 ENTAS STANDARD; PRT; 540 AA.
AC O66198;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOPA.

```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Erwinia.
OX NCBI_TaxID=68334;
RN [1]
RR SEQUENCE FROM N.A.
RC STRAIN:IAM 14479;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -I- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity);
CC -I- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: Belongs to the chaperonin (HSP60) family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or
CC send an email to license@isb-sib.ch).
-----
CC ENBL; AB008153; BAA25239.1; -.
DR HSPF; P05139; IGRN.
DR HAMAP; MF 00600; -.
DR InterPro; IPR01844; Chpnrn Cpn60.
DR InterPro; IPR002423; Gpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1_1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR KEGG; ATP-BINDING; CHAPERONE; ATGP-BINDING.
FT NON TER 540 540
SQ SEQUENCE 540 AA; 56707 MW; B5CD54BDB39373E7 CRC64;
Query Match 7.58; Score 80.5; DB 1; Length 540;
Best Local Similarity 17.7%; Pred.No. 40;
Matches 42; Conservative 46; Mismatches 80; Indels 69; Gaps 9;

QY 13 VASADPIHYDKITEIEINKAIDDA--IAAIQESETIDPMKVPDHRADKPERHVGIVDPKGL 71
Db || :| ::||:::||:::||:::||:::||:::||::|| |::| :
107 VAAGMPEM-----DLKRIGDAVAVADE---LKLSVPCSDSKALAQVGITISANDSA 156

QY 72 AMRNIARGHLKOMRGODANVKGEIVKAHLIGVHD----IVSMEDYLAY ----- 120
Db :| ::||:::||:::||:::||:::||:::||:::||::|| |::| :
157 TVGELLIAAMEKV-----KEGVITVEEGTLQLDELDDVVWGMQFDRIYSFPYFK 207

QY 121 -----KLGDLPHTTHVIDIQDFVFVALSLSEISDSGNLTMTSTFEV 159
Db ||::||::||::||::||::||::||::||::||::||::||::||:| :
208 PETGAVERLESPFILLADKKTSIREMLPFVLVAKAKGKPULLIIABDVGGALLATL----- 262

QY 160 RGFANNVNHSGGL-----SIIDPIFGVLSLDVIITAIFQDTVRKEMTKVLPAFAFKRELEX 212
Db ||::||::||::||::||::||::||::||::||::||::||::||:| :
263 ----VNTWRGIVKVAANKAPGFG---DRRKAMLQDI AVLTCGTVTISEIGMELEK 311

RESULT 63
CH60_ERWHE
ID _CH60_ERWHE STANDARD; PRT; 541 AA.
AC O66216;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOXA OR GROEL.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
```


Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-1- TISSUE SPECIFICITY: Pharyngeal muscle.
-1- MISCELLANEOUS: This protein binds two calcium ions (Potential).
-1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB079299; BAB84566.1; -
EMBL; Z48585; CAA88482.1; -
PIR; T27963; T27963.
HSP; P02588; IPON.
WormPep; ZK673.7; CE01719.
InterPro; IPR003299; Calflagin.
InterPro; IPR002048; EF-hand.
Pfam; PF00036; ehand; 4.
PRINTS; PR01362; CALFLAGIN.
ProDom; PD000012; EF-hand; 2.
SMART; SM00054; Eph; 4.
PROSITE; PS00019; EF_HAND; 2.
Calcium-binding; Repeat.
DOMAIN 28 39 ANCESTRAL CALCIUM SITE 1 (POTENTIAL).
FT CA BIND 64 75 EF-HAND 2 (POTENTIAL).
FT DOMAIN 105 116 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT CA BIND 141 152 EF-HAND 4 (POTENTIAL).
SQ SEQUENCE 160 AA; 18227 MW; 18552E2D8D5D58CE CRC64;

Query Match 7.5%; Score 80; DB 1; Length 160;
Best Local Similarity 23.4%; Pred. No. 11;
Matches 27; Conservative 28; Mismatches 42; Indels 20; Gaps 3;

QY 53 DHADKFERHVGIVDFKGLAMR-----NIEARGLKMKRQGDANVKGE--- 95
Db 15 DQLEQRFKYNMFDKSGKQVIRATQVQILRTWGQAFERDLKLIKEFDAGSGEIEPE 74

QY 96 --EGIVKAHLILGHDDIVSMYDLYKGLDHPHTHV-ISDIQDFVVALSLEISDE 149
Db 75 EFAAMVANFYVNNENDEGLEERLEAFRLYKDEGNGYINVSILRLDILRALDNNVSEE 131

RESULT 66
SYTB_METKA STANDARD; PRT; 590 AA.
AC Q8TX56;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
DE (Phenylalanine--tRNA ligase beta chain) (PheRS).
GN PHE1 OR MK0820.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RA MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaeva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozaykin S.A.;
RT "the complete genome of hyperthermophilic Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By

similarity).
-1- SUBCELLULAR LOCATION: Cytoplasmic.
-1- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain family. Subfamily 2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE010373; AAM02033.1; -
DR HAMAP; MF 00284; -; 1.
DR InterPro; IPR005146; B3_4.
DR InterPro; IPR005147; B5_
DR InterPro; IPR004531; PheT_arch.
DR Pfam; PF03483; B3_4; 1.
DR Pfam; PF03484; B5; 1.
DR TIGRFAMs; TIGR00471; pheT_arch; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 590 AA; 67511 MW; FE1D644780F256BE CRC64;

Query Match 7.5%; Score 80; DB 1; Length 590;
Best Local Similarity 22.6%; Pred. No. 49;
Matches 55; Conservative 28; Mismatches 74; Indels 86; Gaps 12;

QY 20 IYDKITEEINKADDAATAEQSETIDPMKVPD--HADKFERHVGIVDFKGLAMRNI 77
Db 6 IYDKLV-----KILGREVSFEELAHNLIPMLGSDVERIDEREMVETETFFNRPDLISVE 61

QY 78 --ARGLKQM-----KRGDANVKGEIVKA-----HL 103
Db 62 GVARALKGFLGTCIPYNNVRRSDVEARVESLDARPCLAVALVRGVEFEDERDLEHL 121

QY 104 L-----IGVHDDIVSMYDLYKGLDHPHTHV-ISDIQDFVVALSLE 145
Db 122 MEFAQHLHWVGRDKKAAIGH--DFAVEPPLRYFLAD--PNDR-----SWAFE 168

QY 146 ISDEGNITMTSFEV-----RQFANVNNHIGGLSILDPFGLVSLDVLTAIFQDTVRKEM 198
Db 169 PLDPHGEEMTPAEVLRLRHKRGQYAHLV--DGAPILADEBEGVIS-----FPPVINSE 220

QY 199 TKV 201
Db 221 TRV 223

RESULT 67
UNR_RAT STANDARD; PRT; 798 AA.
ID UNR_RAT
AC P18395;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UNR protein.
GN UNR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90370473; PubMed=2204029;
RA Jeffers M., Pacucci R., Pellicer A.;
RT "Characterization of unr; a gene closely linked to N-ras";
RL Nucleic Acids Res. 18:4891-4899(1990).
CC -1- FUNCTION: RNA-binding protein (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the cold-shock domain (CSD) family.

DR SMART; SMO0180; EGF Lam; 3.
DR SMART; SMO0282; LamG; 5.
DR PROSITE; PS00022; EGF 1; UNKNOWN 1.
DR PROSITE; PS01248; LAMININ TYPE EGF; 3.
DR PROSITE; PS50025; LAM G DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 1816
FT DOMAIN 82 131 LAMININ ALPHA-4 CHAIN.
FT DOMAIN 132 186 LAMININ EGF-LIKE 1.
FT DOMAIN 187 240 LAMININ EGF-LIKE 2.
FT DOMAIN 241 255 LAMININ EGF-LIKE 3.
FT DOMAIN 256 825 LAMININ EGF-LIKE 4 (INCOMPLETE).
FT DOMAIN 826 1030 LAMININ II AND I.
FT DOMAIN 1031 1222 LAMININ G-LIKE 1.
FT DOMAIN 1223 1397 LAMININ G-LIKE 2.
FT DOMAIN 1398 1462 LAMININ G-LIKE 3.
FT DOMAIN 1463 1633 LAMININ G-LIKE 4.
FT DOMAIN 1634 1813 LAMININ G-LIKE 5.
FT DOMAIN 1814 523 COILED COIL (POTENTIAL).
FT DOMAIN 524 604 COILED COIL (POTENTIAL).
FT DOMAIN 605 717 COILED COIL (POTENTIAL).
FT DOMAIN 718 799 COILED COIL (POTENTIAL).
FT SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 82 91 BY SIMILARITY.
FT DISULFID 84 98 BY SIMILARITY.
FT DISULFID 101 110 BY SIMILARITY.
FT DISULFID 113 129 BY SIMILARITY.
FT DISULFID 132 146 BY SIMILARITY.
FT DISULFID 157 166 BY SIMILARITY.
FT DISULFID 169 184 BY SIMILARITY.
FT DISULFID 187 202 BY SIMILARITY.
FT DISULFID 189 209 BY SIMILARITY.
FT DISULFID 212 221 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 269 269 INTERCHAIN (PROBABLE).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1088 1088 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1283 1283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1361 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 8 8 C -> S (IN REF. 2).
FT CONFLICT 18 18 C -> Y (IN REF. 2).
FT CONFLICT 248 248 C -> R (IN REF. 3).
FT CONFLICT 297 297 G -> A (IN REF. 3).
FT CONFLICT 431 433 THR -> HPS (IN REF. 2).
FT CONFLICT 679 679 S -> C (IN REF. 3).
FT CONFLICT 703 703 D -> G (IN REF. 2).
FT CONFLICT 706 706 N -> H (IN REF. 2).
FT CONFLICT 728 728 K -> R (IN REF. 2).
FT CONFLICT 730 730 F -> I (IN REF. 2).
FT CONFLICT 779 779 R -> S (IN REF. 1; AA SEQUENCE).
FT CONFLICT 810 810 R -> S (IN REF. 3).
FT CONFLICT 865 867 AEP -> QT (IN REF. 2).
FT CONFLICT 936 936 K -> E (IN REF. 3).
FT CONFLICT 970 970 L -> V (IN REF. 3).
FT CONFLICT 1132 1132 H -> R (IN REF. 2).
FT CONFLICT 1200 1200 F -> I (IN REF. 2).

FT CONFLICT 1382 1382 D -> A (IN REF. 2).
FT CONFLICT 1413 1414 NS -> EP (IN REF. 1).
FT CONFLICT 1489 1489 A -> S (IN REF. 2).
SQ SEQUENCE 1816 AA; 201818 MW; B49C45F3A45999D8 CRC64;
Query Match 7.5%; Score 80; DB 1; Length 1816;
Best Local Similarity 21.3%; Pred. No. 1.8e+02;
Matches 44; Conservative 39; Mismatches 90; Indels 34; Gaps 9;
QY 25 ITEEKALDDAIAIEQSEITIDPMKVDHDKFERHVGIVDFKGLAMRNIEARGLKQM 84
DB 481 LQESINQALDHVRDAEDMNRAT-TPKQDHEKQHEVKEQMEVVGASLSMSADSITIPQL 539
QY 85 K-RQGDANVKGEEGIVKHAHLIGVHDDIVSMYEDLAYKLGDLHPHTH-VISDIQDFVVAL 142
DB 540 TLEELDEIKNASGI-----YAEIDGAKNELOGLKLSNLSNLSHDLVQEAETHAYNL 590
QY 143 -----SLEISDEGNITWTSFEVRO-FANVNVHIG-----GLSILDPIFGVLSD 184
DB 591 QOEADLSRLNLSHSDMGLVQKALDASNVYENIANVSEANETAEALNALNITDRIYDAVSG 650
QY 185 VLTAFQDITVRKMTKVLAPAFKRELE 211
DB 651 IDTQIIYH--KQESDNLNQA--RELQ 673
RESULT 69
FKB4 RABIT STANDARD; PRT; 457 AA.
AC P27124;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE FKS06-binding protein 4 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase) (p59 protein) (HSP binding immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein) (FKBP59).
DE (FKBP59).
CN FKB4 OR P59.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=92165768; PubMed=1537818;
RA Lebeau M.-C., Massol N., Herrick J., Faber L.E., Renoir J.-M., Radanyi C., Baulieu E.-B.;
RA "p59, an hsp 90-binding protein. Cloning and sequencing of its cDNA and preparation of a peptide-directed polyclonal antibody.";
RL J. Biol. Chem. 267:4281-4284(1992).
RN [2]
RP SEQUENCE OF 1-25.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=96154240; PubMed=8579355;
RA Deshpande K.L., Seubert P.H., Tillman D.M., Farkas W.R., Katze J.R.;
RT "Cloning and characterization of cDNA encoding the rabbit tRNA-guanine transglycosylase 60-kilodalton subunit.";
RL Arch. Biochem. Biophys. 326:1-7(1996).
RN [3]
RP DOMAINS.
RX MEDLINE=92335279; PubMed=16311118;
RA Callebaut I., Renoir J.-M., Lebeau M.-C., Massol N., Burny A., Baulieu E.-E., Mornon J.-P.;
RT "An immunophilin that binds M(r) 90,000 heat shock protein: main structural features of a mammalian p59 protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6270-6274(1992).
RN [4]
RP PHOSPHORYLATION BY CK2.
RX MEDLINE=98070781; PubMed=9405642;
RA Miyata Y., Chambrud B., Radanyi C., Leclerc J., Lebeau M.-C., Renoir J.-M., Shirai R., Catsall M.-G., Yohara I., Baulieu E.-B.;
RT "Phosphorylation of the immunosuppressant FK506-binding protein FKBP52

by casein kinase II: regulation of HSP90-binding activity of FKBP52.";
Proc. Natl. Acad. Sci. U.S.A. 94:14500-14505(1997).
[5]
INTERACTION WITH NR3C1; DYNEIN AND HSP90.
MEDLINE=20069677; PubMed=10601253;
Silverstein A.M., Galligniana M.D., Kanelakis K.C., Radanyi C.,
Renoir J.-M., Pratt W.B.,
"Different regions of the immunophilin FKBP52 determine its
association with the glucocorticoid receptor, hsp90, and cytoplasmic
dynein.";
J. Biol. Chem. 274:36980-36985(1999).
[6]
INTERACTION WITH PHVH.
MEDLINE=99162565; PubMed=10051602;
Chambraud B., Radanyi C., Camonis J.H., Rajkowski K., Schumacher M.,
Baulieu E.-E.,
"Immunophilins, refsum disease, and lupus nephritis: the peroxisomal
enzyme phytanoyl-CoA alpha-hydroxylase is a new FKBP-associated
protein.";
Proc. Natl. Acad. Sci. U.S.A. 96:2104-2109(1999).
[7]
STRUCTURE BY NMR OF 1-148.
MEDLINE=96374215; PubMed=8780506;
Craescu C.T., Rouviere N., Popescu A., Cerpolini E., Lebeau M.-C.,
Baulieu E.-E., Mispelster J.,
"Three-dimensional structure of the immunophilin-like domain of
FKBP59 in solution.";
Biochemistry 35:11045-11052(1996).
CC -!- FUNCTION: Component of unactivated mammalian steroid receptor
complexes that sediment at 8-10 S. May have a rotamase activity.
May play a role in the intracellular trafficking of hetero-
oligomeric forms of steroid hormone receptors.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
(omega=0).
CC -!- SUBUNIT: Associates with HSP90 and HSP70 in unactivated steroid
hormone receptor complexes. Also interacts with peroxisomal
phytanoyl-CoA alpha-hydroxylase (PHVH) (By similarity). Interacts
with NR3C1 and dynein.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- PTM: Phosphorylation by CK2 results in loss of HSP90 binding
activity.
CC -!- SIMILARITY: Belongs to the FKBP-type PPIase family.
CC -!- SIMILARITY: Contains 2 FKBP-type PPIase domains.
CC -!- SIMILARITY: Contains 3 TPR repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M84474; AAA31438.1; -
CC EMBL; M84988; AAA31439.1; -
CC PIR; A42386; A42386
CC PDB; 1ROT; 07-DEC-96.
CC PDB; 1ROU; 07-DEC-96.
CC InterPro; IPR001179; FKBP PPIase.
CC InterPro; IPR008941; TPR-like.
CC InterPro; IPR001440; TPR.
CC Pfam; PF00254; FKBP; 2.
CC Pfam; PF00515; TPR; 3.
CC PROSITE; PS00453; FKBP PPIASE 1; 1.
CC PROSITE; PS00454; FKBP PPIASE 2; 2.
CC PROSITE; PS00509; FKBP PPIASE 3; 2.
CC K01 Isomerase; Rotamase; TPR repeat; Repeat; Nuclear protein;
3D-structure; Phosphorylation.
CC INIT MET 0
CC DOMAIN 49 137 PPIASE, FKBP-TYPE 1.
CC DOMAIN 166 282 PPIASE, FKBP-TYPE 2.
CC REPEAT 269 302 TPR 1.
CC REPEAT 318 351 TPR 2.
CC FT

TPR 3.
PHOSPHORYLATION (BY CK2).
S -> H (IN REF. 3).
EG -> FI (IN REF. 3).
S -> T (IN REF. 3).
385
353 REPEAT
142 MOD RES
14 CONFLICT
20 CONFLICT
25 CONFLICT
23 TURN
26 TURN
32 STRAND
38 TURN
48 TURN
51 STRAND
61 TURN
64 STRAND
69 TURN
72 STRAND
76 TURN
80 HELIX
87 TURN
95 TURN
98 STRAND
101 STRAND
108 TURN
111 TURN
115 TURN
118 TURN
120 STRAND
127 STRAND
457 AA; 51344 MW; EC58CC4BCF66A44A CRC64;
SQ
Query Match 7.4%; Score 79.5; DB 1; Length 457;
Best Local Similarity 22.1%; Pred. No. 40;
Matches 27; Conservative 29; Mismatches 49; Indels 17; Gaps 3;
QY 18 DPHYDKITEINKAIDDAIAEQSETI-----DPMKVPDHD-KFERHV 62
Db 74 DKSFDLKGKGVKAWDAVATMKVGLCRITCKPEYAYGSGSPKIPPNATLVFB-V 131
QY 63 GIVDPKGLAMNIEARGLKQKMGDANVGEGIVKAHLIGVHDIVSMEDLAYKL 122
Db 132 ELFPKGEDLTDEDDGGIRIRTRGEGYARNPDGAIVEALEGGYKDRFDQRELFV 191
QY 123 GD 124
Db 192 GE 193
RESULT 70
CH60 ERWCA STANDARD; PRT; 540 AA.
ID CH60 ERWCA STANDARD; PRT; 540 AA.
AC O86220;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOXA OR GROEL.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM 12633;
RA Harada H., Ishikawa H.,
RT "Phylogenetical relationship based on groE genes among phenotypically
related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
species.";
RT J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

CC -----
DR ENBL; AB008152; BAA25237.1; -.
DR HSP; P06139; IGEL.
DR HAMAP; MF_00600; -, 1.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR PROSITE; PS00304; TCOMPLEXTCP1.
KW Chaperone; ATP-binding.
FT NON TER 540 540
SQ SEQUENCE 540 AA: 56292 MW: 7E41FEA699566719 CRC64:

```

Microb. Pathog. 17:271-275(1994).

```

[3] SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd I., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[4]
RC SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=92531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland W., Kodyranyi V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (by similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (by similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@sb-sib.ch).

```

CC	ENBL; ASO08903; AAL23153.1; -	DR	SVGene; SG32???; grol.	DR	547 AA; 57154 MW;
DR	ENBL; U01039; AAR53277.1; -	DR	HAMA; NF_00600; -; 1.	DR	SEQUENCE
DR	ENBL; ALU6283; CAD06810.1; -	DR	InterPro; IPR001844; Chaprpn1	DR	
DR	ENBL; AS015649; AAO71833.1; -	DR	InterPro; IPR008950; GroEL-ATP	DR	
DR	HSP; P06139; 1KID.	DR	Pfam; PR00118; cpn60 TCP1.1.	DR	
DR		DR	PRINTS; P00238; CPN60NTN160.	DR	
DR		DR	PRINTS; P003304; TCOMPLEXTCP1.	DR	
DR		DR	PROSITE; PS00296; CHAPERONIN60.	DR	
KW	Chaperone; ATP-binding; Comple	DR	Chaperone; ATP-binding; Comple	DR	
FT	INIT MET 0	DR	INIT MET 0	DR	
SQ	SEQUENCE	DR	SEQUENCE	DR	

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
RT LT2,"
RL Nature 413:852-856(2001).

121
RP SEQUENCE FROM N.A.
RC SPECIES-S.typhi; STRAIN-Ty2 / ATCC 700931;
RX MEDLINE=95231266; PubMed=7715425;
RA Lindler L.E., Hayes J.M.;
RT "Nucleotide sequence of the *Salmonella typhi* groEL heat shock gene.";

Query Match 7.4%; Score 79.5; DB 1; Length 819;
Best Local Similarity 21.1%; Pred. No. 79;


```

Query Match      7.4%; Score 79.5; DB 1; Length 906;
Best Local Similarity 19.4%; Pred. No. 89;
Matches 48; Conservative 47; Mismatches 82; Indels 71; Gaps 10;

QY      18 DPIH--YOKITE-----EINKAIDDAIA-----AIEOSETIDPMKVDPHAD 56
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      361 DELNIAIDKMTIKTDLRQLRKAMVDHISDSPLETNVPLLVIEAASGNEKEVYQAQ 420
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY      57 KPERHVG-IVDFKELANRNTEARGLKQMK-----RQ 87
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      421 VPFEHANKLIVEANLACSISSNEEGVKLVMAATQIDSLCPQVINAALTAAARPOSKVAQ 480
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY      88 GPANYKGBEGIVKAKHLITGVHDDIVSMEYDLAYKLGDLPHPTTHVDSIQDFVVALSLRIS 147
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      481 DNMDVFQDQWEKQVRVLTAEAVDDITSVDDELFSVS-----ENHILEDNVKCVIA--LQEG 532
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY      148 DEGNITMTSPFEVQFQPNVANHIGLSLIDRIFGVLSDLVLTAFQDTPVRKMTKVLA----- 203
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      533 DVDVTDRTAGAIRGRARVHI-----INAMENYETGVYTEKV-LEATKLLSETVM 583
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY      204 PAFKSELE 211
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      584 PRFAEQVE 591
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

```

RESULT 74	ARGR_CLOTE	STANDARD;	PRT;	153 AA.
ID	ARGR_CLOTE			
AC	Q894H3;			
DT	15-MAR-2004 (Rel. 43, Created)			
DT	15-MAR-2004 (Rel. 43, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Arginine repressor.			
GN	ARGR OR C1C01572.			
OS	Clostridium tetani.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1513;			
[1]	__TaxID=1513;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=Massachusetts / E88;			
RC	MEDLINE=22457253; PubMed=12552129;			
RX	Bruggemann H., Baeumer S., Fricke W.F., Wierse A., Liesegang H.,			
RA	Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,			
RA	Gotteschalk G.;			
RT	"The genome sequence of Clostridium tetani, the causative agent of			
RT	tetanus disease."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).			
CC	!- FUNCTION: Regulates arginine biosynthesis genes (By similarity).			
CC	!- PATHWAY: Arginine biosynthesis; regulation.			
CC	!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).			
CC	!- SIMILARITY: Belongs to the argR family.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC				
CC	EMBL; AE015941; AAC036119.1; -.			
CC	HAMAP; MF_00173; -; 1.			
DR	InterPro; IPR001669; Arg repress.			
DR	InterPro; IPR009058; Wing hlx DNA_bnd.			
DR	Pfam; PF01316; Arg repressor; 1			
DR	Pfam; PF02863; Arg_repressor C; 1.			
DR	PRINTS; PR01467; ARGREPRESSOR.			
DR	Transcription regulation; DNA-binding; Trans-acting factor; Repressor;			
KW	Arginine biosynthesis; Complete proteome.			
SQ	SEQUENCE 153 AA; 17120 MW; 27694F52058CF26 CRC64;			

```

Query Match          7.4%; Score 79; DB 1; Length 153;
Best Local Similarity 24.7%; Pred. No. 12;
Matches 42; Conservative 30; Mismatches 60; Indels 38; Gaps 9

Qy 49 MKVPDHADKFERHVGIVDFKG-----ELAMRNIE-----ARGLKMKRQGDANVK 93
Db 1 MKVTRH---EKILELIERKDIOEELAEELRKSGETIQTATVSRDIKELKL---IKVL 53

Qy 94 GEGIVKXAHLLIGVHDDIVSMEYDLAKVIGDLHPHTHV-ISDIQDFVVALSLETDEGNI 152
Db 54 GEQSKYK-----YAAIVKNENLSDKLANIPSHSVSVSNINNFVVVKTL--SGSNA 104

Qy 153 TMTSFVQRQFANVVNHIGGISLPIFGVLSDLVLTAFQDTPVRKEMTKVL 202
Db 105 AAEASIDLSNKEFIAGTIAG---DNTIFIMARTSEQAFE-IVKKMKRVI 148

RESULT 75
KPYK_THELI
AC_KPYK_THELI STANDARD; PRT; 220 AA.
ID Q56301;
DR 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pyruvate kinase (EC 2.7.1.40) (PK)(Fragment).

```

RESULT 75

```

KPYK_THELI
ID STANDARD; PET; 220 AA.
AC Q56301;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pyruvate kinase (EC 2.7.1.40) (PK) (Fragment).
GN PKI.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=2265;
RN [1]
R1 SEQUENCE FROM N.A.
R2 STRAIN=DSM 5473;
R3 MEDLINE=95370154; PubMed=7642512;
R4 Kletzin A., Mukund S., Kelley-Crouse T.L., Chan M.K., Rees D.C.,
R5 Adams M.W.W.;
R6 "Molecular characterization of the genes encoding the tungsten-
R7 containing aldehyde ferredoxin oxidoreductase from Pyrococcus
R8 furiosus and formaldehyde ferredoxin oxidoreductase from Thermococcus
R9 litoralis.";
R10 J. Bacteriol. 177:4817-4819 (1995).
R11
R12 C- CATALYTIC ACTIVITY: ATP + pyruvate = ADP + phosphoenolpyruvate.
R13
R14 SUBUNIT: Homotetramer (By similarity).
R15
R16 SIMILARITY: Belongs to the pyruvate kinase family.
R17
R18
R19
R20
R21
R22
R23
R24
R25
R26
R27
R28
R29
R30
R31
R32
R33
R34
R35
R36
R37
R38
R39
R40
R41
R42
R43
R44
R45
R46
R47
R48
R49
R50
R51
R52
R53
R54
R55
R56
R57
R58
R59
R60
R61
R62
R63
R64
R65
R66
R67
R68
R69
R70
R71
R72
R73
R74
R75
R76
R77
R78
R79
R80
R81
R82
R83
R84
R85
R86
R87
R88
R89
R90
R91
R92
R93
R94
R95
R96
R97
R98
R99
R100
R101
R102
R103
R104
R105
R106
R107
R108
R109
R110
R111
R112
R113
R114
R115
R116
R117
R118
R119
R120
R121
R122
R123
R124
R125
R126
R127
R128
R129
R130
R131
R132
R133
R134
R135
R136
R137
R138
R139
R140
R141
R142
R143
R144
R145
R146
R147
R148
R149
R150
R151
R152
R153
R154
R155
R156
R157
R158
R159
R160
R161
R162
R163
R164
R165
R166
R167
R168
R169
R170
R171
R172
R173
R174
R175
R176
R177
R178
R179
R180
R181
R182
R183
R184
R185
R186
R187
R188
R189
R190
R191
R192
R193
R194
R195
R196
R197
R198
R199
R200
R201
R202
R203
R204
R205
R206
R207
R208
R209
R210
R211
R212
R213
R214
R215
R216
R217
R218
R219
R220
R221
R222
R223
R224
R225
R226
R227
R228
R229
R230
R231
R232
R233
R234
R235
R236
R237
R238
R239
R240
R241
R242
R243
R244
R245
R246
R247
R248
R249
R250
R251
R252
R253
R254
R255
R256
R257
R258
R259
R260
R261
R262
R263
R264
R265
R266
R267
R268
R269
R270
R271
R272
R273
R274
R275
R276
R277
R278
R279
R280
R281
R282
R283
R284
R285
R286
R287
R288
R289
R290
R291
R292
R293
R294
R295
R296
R297
R298
R299
R300
R301
R302
R303
R304
R305
R306
R307
R308
R309
R310
R311
R312
R313
R314
R315
R316
R317
R318
R319
R320
R321
R322
R323
R324
R325
R326
R327
R328
R329
R330
R331
R332
R333
R334
R335
R336
R337
R338
R339
R340
R341
R342
R343
R344
R345
R346
R347
R348
R349
R350
R351
R352
R353
R354
R355
R356
R357
R358
R359
R360
R361
R362
R363
R364
R365
R366
R367
R368
R369
R370
R371
R372
R373
R374
R375
R376
R377
R378
R379
R380
R381
R382
R383
R384
R385
R386
R387
R388
R389
R390
R391
R392
R393
R394
R395
R396
R397
R398
R399
R400
R401
R402
R403
R404
R405
R406
R407
R408
R409
R410
R411
R412
R413
R414
R415
R416
R417
R418
R419
R420
R421
R422
R423
R424
R425
R426
R427
R428
R429
R430
R431
R432
R433
R434
R435
R436
R437
R438
R439
R440
R441
R442
R443
R444
R445
R446
R447
R448
R449
R450
R451
R452
R453
R454
R455
R456
R457
R458
R459
R460
R461
R462
R463
R464
R465
R466
R467
R468
R469
R470
R471
R472
R473
R474
R475
R476
R477
R478
R479
R480
R481
R482
R483
R484
R485
R486
R487
R488
R489
R490
R491
R492
R493
R494
R495
R496
R497
R498
R499
R500
R501
R502
R503
R504
R505
R506
R507
R508
R509
R510
R511
R512
R513
R514
R515
R516
R517
R518
R519
R520
R521
R522
R523
R524
R525
R526
R527
R528
R529
R530
R531
R532
R533
R534
R535
R536
R537
R538
R539
R540
R541
R542
R543
R544
R545
R546
R547
R548
R549
R550
R551
R552
R553
R554
R555
R556
R557
R558
R559
R560
R561
R562
R563
R564
R565
R566
R567
R568
R569
R570
R571
R572
R573
R574
R575
R576
R577
R578
R579
R580
R581
R582
R583
R584
R585
R586
R587
R588
R589
R590
R591
R592
R593
R594
R595
R596
R597
R598
R599
R600
R601
R602
R603
R604
R605
R606
R607
R608
R609
R610
R611
R612
R613
R614
R615
R616
R617
R618
R619
R620
R621
R622
R623
R624
R625
R626
R627
R628
R629
R630
R631
R632
R633
R634
R635
R636
R637
R638
R639
R640
R641
R642
R643
R644
R645
R646
R647
R648
R649
R650
R651
R652
R653
R654
R655
R656
R657
R658
R659
R660
R661
R662
R663
R664
R665
R666
R667
R668
R669
R670
R671
R672
R673
R674
R675
R676
R677
R678
R679
R680
R681
R682
R683
R684
R685
R686
R687
R688
R689
R690
R691
R692
R693
R694
R695
R696
R697
R698
R699
R700
R701
R702
R703
R704
R705
R706
R707
R708
R709
R710
R711
R712
R713
R714
R715
R716
R717
R718
R719
R720
R721
R722
R723
R724
R725
R726
R727
R728
R729
R730
R731
R732
R733
R734
R735
R736
R737
R738
R739
R740
R741
R742
R743
R744
R745
R746
R747
R748
R749
R750
R751
R752
R753
R754
R755
R756
R757
R758
R759
R760
R761
R762
R763
R764
```

```

Query Match      7.4%; Score 79; DB 1; Length 220;
Best Local Similarity 23.0%; Pred. No. 13;
Matches 48; Conservative 36; Mismatches 71; Indels 54; Gaps 10;

Qy 19 PIHYDKI-----TBEINKAIDD--AIAAI-----EQSETIDMKVPQHADKF 58
Db 8 PSHKTKIIATIGPASKOKESIKMKIKAGMSVARINFSHGTLSEHAKTITVLR--DVAEKI 65

```

QY 59 ERHVGVDFKGLAMRNIEARGIKQMKROG-----ANVKGEGIVKAHL-----LIGV 107
 KW Ribosomal protein.
 SQ SEQUENCE 406 AA; 44702 MW; A56AFA6CEAF291F6 CRC64;
 DB 66 ERRVAILGDLPGKRGKRGKSVTLRKGDGVKLVLTTRDIEGDETTIPVEFKDLPKLVSK 125
 QY 108 HDDIVSMEYDLAYKGLDHPHTTHWISDIQDFVVALSLEISDEG-----NITMTS 156
 DB 126 GDTIYLSGDIYMLRVEVRE-----NEVECVVNGGILFSGHKGINIPKANLPLEAITPRD 180
 QY 157 FEVRQFANVNVH-----IGGLSILDIPIFOVL 182
 DB 181 FEIIEFA--IEHGVDAIGLSFVGSVDVL 207
 RESULT 76
 RL4B_ARATH STANDARD; PRT; 406 AA.
 ID RL4B_ARATH
 AC Q9SF40;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L4-2 (L1).
 GN RPL4B OR AT3G09630 OR F118.22.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurasids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ev. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger H., Ansoerge W., Unseld M.,
 RA Fattmann B., Valle G., Bloecher M., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choise N., Artiguenave P., Robert C., Brottier P.,
 RA Wincker P., Catolico L., Weissenbach J., Sauvin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Dronek H., Erle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonet G.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordisiek G.,
 RA Reichelt J., Schaefer M., Schoen O., Bagues M., Terol J., Clement J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy B.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Manfauit G., Argirou A., Flores M., Liguori R., Vitale D.,
 RA Monfort A., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Millscher J., Sellers P., Gill J.B., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Igesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shimpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana."
 RL Nature 408:820-822 (2000).
 CC -!- SIMILARITY: Belongs to the L4E family of ribosomal proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AC016661; AAF23293.1;
 CC InterPro; IPR002136; Ribosomal_L4/L1E.
 CC Pfam; PF00573; Ribosomal_L4; 1.
 CC

DR PROSITE; PS00939; RIBOSOMAL_L1E; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 406 AA; 44702 MW; A56AFA6CEAF291F6 CRC64;
 Query Match 7.4%; Score 79; DB 1; Length 406;
 Best Local Similarity 26.4%; Pred. No. 38;
 Matches 29; Conservative 24; Mismatches 41; Indels 16; Gaps 4;
 QY 1 MKKELLIAAFAVAVSADPI---HYDKITEINKAIDATAAIBQS-----ETIDPMKVPD 53
 DB 124 MKRHAIVSAIAATAVPALVMARGHKIENVPEMPLVSDSAENVEKTSAAIKVLKQIGAYD 183
 QY 54 HADKFERHVGIVDFKGLAMRNIEARGIKQMKROGDANVKGEG--IVKA 101
 DB 184 DAERAKNSIGIRPGKGM-----RNRVVISRKGLVPLVVTGEGSKIYKA 226
 RESULT 77
 LE11_PYRAB STANDARD; PRT; 486 AA.
 ID LE11_PYRAB
 AC Q9UZ08;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 2-isopropylmalate synthase 1 (EC 2.3.3.13) (Alpha-isopropylmalate
 synthase 1) (Alpha-IPM synthetase 1).
 GN LEUAI OR PYRAB13490 OR PAB0890.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OC NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GES / Orsay;
 RX MEDLINE=22511545; PubMed=12622808;
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
 RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
 RA Van der Oost J., Weissenbach J., Zivanovic V., Forterre P.;
 RT "An integrated analysis of the genome of the hyperthermophilic
 archaeon Pyrococcus abyssi."
 RL Mol. Microbiol. 47:1495-1512 (2003).
 CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
 CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
 CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O =
 CC 2-hydroxy-2-isopropylsuccinate + CoA, first step.
 CC -!- PATHWAY: Leucine biosynthesis; first step.
 CC -!- SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate
 CC synthase family. LeuA 1 subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ248287; CAB50254.1;
 CC FIR; A75045; A75045.
 CC HAMAP; MF_01025; 1.
 CC InterPro; IPR002034; AIPW/Hcit synth.
 CC InterPro; IPR000891; HMGL-like.
 CC InterPro; IPR005671; LeuA_bact_synth.
 CC Pfam; PF00682; HMGL-like; 1.
 CC TIGRfam; TIGR00973; leuA_bact; 1.
 CC PROSITE; PS00815; AIPW_HOMOCIT_SYNTH_1; 1.
 CC PROSITE; PS00816; AIPW_HOMOCIT_SYNTH_2; 1.
 CC Leucine biosynthesis; Transferrase; Complete proteome.
 CC SEQUENCE 486 AA; 53768 MW; 721D5B4724CC49B2 CRC64;
 Query Match 7.4%; Score 79; DB 1; Length 486;
 Best Local Similarity 21.0%; Pred. No. 47;
 DR

DR EMBL; AE010580; AAL94893.1; -.
DR HAMAP; MF_00036; -, 1
DR InterPro; IPR003156; DHAL.
DR InterPro; IPR002318; tRNA-synt_2c.
DR InterPro; IPR006193; tRNA-synt_Ala.
DR Pfam; PF02272; DHAL; 1.
DR Pfam; PF01411; tRNA-synt_2c; 1.
DR PRINTS; PR00980; TENASYNTHALA.
DR PRINTS; TIGR00344; alas; 1.
DR PROSITE; PS00860; AA_tRNA_LIGASE_II_ALA; 1.
DR K01497; K01497; Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.

CC EMBL; Y15269; CAA75558.1; -
DR PIR; T07825; T07825.
DR GO; GO:0009535; C:thylakoid membrane (sensu Viridiplantae); IEP.
DR GO; GO:0006950; P:response to stress; IEP.
DR InterPro; IPR006843; PAP fibrillin.
DR Pfam; PF04755; PAP fibrillin; 1.
KW Chloroplast; Thylakoid; Membrane; Transit peptide.
FT TRANSIT 1 63 CHLOROPLAST.
FT CHAIN 64 326 LIGHT-INDUCED PROTEIN.
FT CONFLICT 286 286 T -> I (IN REF. 2).
FT SEQUENCE 326 AA; 35635 MW; 1984482198724DEB CRC64;
Query Match 7.4%; Score 78.5; DB 1; Length 326;
Best Local Similarity 25.4%; Pred. No. 33;
Matches 32; Conservative 21; Mismatches 44; Indels 29; Gaps 5;
QY 27 EEINKATDDAIAAEQSEITIDPMKVPDPHADKFERHVGIVDFKGLAMRNIEARGLKMKR 86
Db 179 EELSTQDSSFTVONS-----VFAGPLATTSISTNAKEVRS 217
QY 87 QGDANVKGEGIVKAHLIGVHDIV---SMEYDLYAKGLDLHPTTHVISDIQDFVALS 143
Db 218 PKRQIKFEGIICTPQLT---DSIVLPENVEF-LGQKI-DLSFFKGLITSVDQTASSVA 272
QY 144 LEISDE 149
Db 273 KSISSQ 278
RESULT 80
YE72_HA2IN STANDARD; PRT; 351 AA.
AC P44206;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein H11472 precursor.
GN H11472
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

CC or send an email to license@isb-sib.ch.
CC EMBL; U32825; AAC23120.1; -
DR PIR; I64030; I64030.
DR TIGR; H11472; -
DR InterPro; IPR002491; Peripla BP.
DR Pfam; PF01497; Peripla BP 2; 1.
KW Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 351 PROTEIN H11472.
FT SEQUENCE 351 AA; 39205 MW; 166842B3F4D2BBED CRC64;
Query Match 7.4%; Score 78.5; DB 1; Length 351;
Best Local Similarity 26.3%; Pred. No. 35;
Matches 44; Conservative 36; Mismatches 54; Indels 33; Gaps 11;
QY 2 MKFLIIA-----AVAFVAVSADPIHYKITBEINKAIDAI--AAEQSEITIDPMKVPDHA 55
Db 3 LKSLIIACLLSLSSLSFSAI-ADRIITDQDRKV--TIPDHINRAVVLQHOTLN---IAVOL 56
QY 56 DKFERHVGIV-DFKGLAMRNIE-ARGIKOMKROGDANVKGEGIVKAHLIGVHDIVS 113
Db 57 DATKQIVGVLSNWKQKQKGNVRLAPELENMMPGDLN-----SVNIESLLAKPDPVF 110
QY 114 MEYDLYAKGLDLHPTTHV--ISDIQDFVALSL---BISDEGNITMT 155
Db 111 VT-----NYAPSEMIKQISDVNIPVAISLRTGVEGKGLNPT 149
RESULT 81
LE12_METTH STANDARD; PRT; 505 AA.
AC Q27525;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 2-isopropylmalate synthase 2 (EC 2.3.3.13) (Alpha-isopropylmalate
DE synthase 2) (Alpha-IPM synthetase 2).
GN LEUA2 OR MTH1481.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OC NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O =
CC 2-hydroxy-2-isopropylsuccinate + CoA.
CC -!- PATHWAY: Leucine biosynthesis; first step.
CC -!- SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate
CC synthase family. Leua 1 subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

```

CC EMBL; AB000909; AAB85956.1; -.
DR PIR; D69064; D69064.
DR HAMAP; MF_01025; -. 1.
DR InterPro; IPR002034; AIPM/HcIt synth.
DR DR Pfam; PF000891; HMGL-like.
DR DR Pfam; PF00682; HMGL-like; 1.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
DR Leucine biosynthesis; Transfrase; Complete proteome.
DR KW SEQUENCE 505 AA; 54674 MW; 7DEDE80E144C5DA5 CRC64;
SQ
Query Match 7.4%; Score 78.5; DB 1; Length 505;
Best Local Similarity 20.6%; Pred No. 54;
Matches 40; Conservative 36; Mismatches 73; Indels 45; Gaps 8
QY 30 NKAI--DDAIA-----AIQSEITDPM--KVPDHADK--FERHVGIVDFKGLAMR 74
Db 284 NNAIVGENAFAHEAGIHVHGLEAKETVEPITPEMVGHKRRIVLGHKGTGANALRSKLQ 343
QY 75 NTEARG-----LKOMKEQGDANV-----KBEIGVKAHLILGVHDDIV 112
Db 344 GEMKEEQFTYGVKVLGDKRKRIITDRLAMAVTILGKAREIVKLEGVAMTGESV 403
QY 113 SMEDYLAYLKLGLDHPHTTHV--ISDIQDFVWALSLEISDEGNITMTSFEVRQFANVNHIG 170
Db 404 MPTATVCLRIGDEVXTTSMTGVPQDAINAIQSLVSETADIELDEVNIEAIT-----G 457
QY 171 GLSILDPFPGVLS D 184
Db 458 GTNALAEVFEVMSD 471
RESULT 82
YAC8 YEAST
ID YAC8 YEAST STANDARD; PRT; 528 AA.
AC P39734; P39733;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 58.8 kDa protein in MYO4-DES2 intergenic region.
GN YAL028W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=55249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J.J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RA "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813 (1995).
RN [2]
RP REVISIONS.
RA Vo D.T.;
RC Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC
DB EMBL; U12980; AAC05004.1; -.
DR PIR; S70295; S70295.
DR GERMOnline; 138370; -.
DR SGD; S0000026; YAL028W.

```

```

Matches 44; Conservative 43; Mismatches 91; Indels 49; Gaps 7;
QY 13 VASADPIHYDKITEEINKAIDDAIAEQSETIDPMKVPDPHADKFERHVGIVDFKGLA 72
DB 107 VAAGNMPX---DLKRGIDKAVDSNV-----EKLRAQAQCSKESITQVGSISANSRA 157
QY 73 MRNTEARGLKQMKRQGDANVKGEGIVKAHLIG-----VHDDIVSMEYDL 118
DB 158 IGDIIAEMKVGKNGVITVEEGQGLSNELSVVEGMOFDRGYLSPFYITNQDSCAVELDN 217
QY 119 AY-----KLGDLHPHTHVSIDQFVVALSLEISDEGNITMTSFEVQRQANVVNHI 169
DB 218 PYILLVKKVSSIRELLPVLESVAKASRLIIAEDIDGEGALATL-----VNNL 267
QY 170 GGL-----SILDPIFGVLSVLTAFQDVTVRKMTKVLAPAFKRELEK 212
DB 268 RGVRAAAVAPGPG---DNRKAMLEIAVLTAGTVISEBIGLELEK 311

RESULT 84
CH60_XANCH STANDARD; PRT; 546 AA.
AC Q8RIT7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL.
OS Xanthomonas campestris (pv. phaseoli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=29445;
RN [1]
RP SEQUENCE FROM N.A.
RA Hsu C.-C., Yu Y.-J., Yang M.-T.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF426387; AAL74150.1; -
CC HAMAP; MF_00600; -; 1
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PROSITE; PS00304; TCOMPLEXTCP1.
CC Chaperone; ATP-binding.
CC SEQUENCE 546 AA; 57225 MW; 9DDAB51DED42BA5C CRC64;

Query Match
Best Local Similarity 7.4%; Score 78.5; DB 1; Length 546;
Matches 49; Conservative 41; Mismatches 79; Indels 71; Gaps 11;
QY 13 VASADPIHYDKITEEINKAIDDAI-RAEQSETIDPMKVPDPHADKFERHVGIVDFKGL 71
DB 107 VAAGNMPX-----DLKRGIDQAVKAAVVELKNISK---PTTDDKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKQMKRQGDANVKGEGIVKAHLIGVH-----DDIVSMEYDLAY 120

```

```

DB 157 SIGNIIAEMQKV-----GKEGVITVEGSGLENELDVVEGMOFDRGYLSPFYINN 207
QY 121 -----KLGDLHPHTH--VISDIOFVVAL-----SLEISDEGNITMTSFEV 159
DB 208 QOSQASLDLDDPFIILDKKISNVRDLPLVLEGVAKAGKAPLLIVAEVEGEGALATL----- 262
QY 160 RQFANVVNHIGL-----SILDPIFG-----VLSVLTAFQDVTVRKMTKVLAPAFKREL 210
DB 263 -----VNTIRGIVKVAVKAPGFGDRRKAMLEDMVAVLGTGTVISERVGLAEKATIKDL 317

RESULT 85
CH60_COLMA STANDARD; PRT; 549 AA.
AC Q93GT8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL.
OS Colwellia maris (Vibrio sp. (strain ABE-1)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Colwellia.
OX NCBI_TaxID=77524;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22883852; PubMed=12898133;
RA Yamauchi S., Okuyama H., Morita E.H., Hayashi H.;
RT "Gene structure and transcriptional regulation specific to the groESL
RT operon from the psychrophilic bacterium Colwellia maris.";
RL Arch. Microbiol. 180:272-278(2003).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB073221; BAB70476.2; -
CC HAMAP; MF_00600; -; 1
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60 TCP1; 1
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC SEQUENCE 549 AA; 57690 MW; 3A194880966062A2 CRC64;

Query Match
Best Local Similarity 7.4%; Score 78.5; DB 1; Length 549;
Matches 52; Conservative 36; Mismatches 82; Indels 65; Gaps 11;
QY 13 VASADPIHYDKITEEINKAIDDAIAEQSETIDPMKVPDPHADKFERHVGIVDFKGLA 72
DB 107 IAAGNMPX---DLKRGIDKAVIAAVALKESST-----PVTDNKAIEQVGTISANSDET 157
QY 73 MRNIEARGLKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKLGDLHP----- 127
DB 158 VGIITATANEKVGEGVITV--BEG-----QALTELDVVEGMOFDRGY-----LSFYINK 207
QY 128 -----TTHVISDIOFVVALS-----LEISD--EGNITMTSFEVRQ 161

```



```
Db 208 QENGTVLENPPFILLVKKISNIRELLTLEGVAKAGKELLIIAEDVEGEALATL-----262
Qy 162 FANVNHIGL-----SILDPFGVLSVLTATFQDTRKEMTKVLAPAFKRELEK 212
Db 263 ---VNNMRGIVKVAAPGFG---DRRKAMLDQVATLTAGTVISEIGMELEK 311

RESULT 86
LON CHLMU STANDARD; PRT; 819 AA.
AC QPKS6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR TC0623
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback I., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: Degrades short-lived regulatory and abnormal proteins in
CC presence of ATP. Hydrolyzes two ATPs for each peptide bond cleaved
CC in the protein substrate (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002331; AAF39454.1; -.
CC PIR; B81681; B81681.
CC MEROPS; S16.002; -.
CC TIGR; TC0623; -.
CC
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR001270; Chaprinin_c1pA/B.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_Lon.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008268; Peptid_S16_AS.
CC InterPro; IPR001984; Peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00300; CLPPROTEASEA.
CC PRINTS; PR00830; ENDOLAPTAASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC TIGRFAMS; TIGR00763; Lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
CC Hydrolase; Serine protease; ATP-binding; Complete proteome.
KW NP_BIND 392 399
FT ACT_SITE 724 724
FT BY SIMILARITY.
```

```
SQ SEQUENCE 819 AA; 91843 MW; DD87D5D701E1B063 CRC64;
Query Match 7.4%; Score 78.5; DB 1; Length 819;
Best local similarity 21.8%; Pred. No. 95;
Matches 51; Conservative 38; Mismatches 78; Indels 67; Gaps 11;

Qy 16 SADPIHYDKITEINKAIDDAIAAEQSETIDP-----MKVP-----52
Db 6 NTDQNLDPNASEVEKLLDSEAEAEKTDHPTPPSELFILPNKRPPFGMAAPLLIEAG 65
Qy 53 DH-----ADKPERHYGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEGIVKAHLII 105
Db 66 PHYEVLTLLAKSSQKHGLVLTKKEDA--NTLKIGFNLHRVG-----VSARIL--112
Qy 106 GVHDDIVSMYEDLAYKLGDLHPHTTHVVISDIQDFVALSLSLEISDEGNITMTSFVRQPA-N 164
Db 113 ---RIMPIEGGSAQVLLSIEDIRIVKVPQDKYLKAKVAYHKE-NKELTE-ELKAYSIS 166
Qy 165 VVNHIGLSILDPF-----GVLSQVLTATFQDTRKEMTKVL 202
Db 167 IVSIKDLLKLNLPLFKEELQIFLGHSDFTEPGKLADFSVALTTAT-REELQEV 219

RESULT 87
SYA METJA STANDARD; PRT; 892 AA.
AC Q57984;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
GN ALAS OR MJO564.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=6880887;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geodhagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RT Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67505; AAB98556.1; -.
CC PIR; D64370; D64370.
CC TIGR; MJO564; -.
CC HAMAP; MF_00036; -.
CC InterPro; IPR003156; DHHA1.
CC InterPro; IPR002318; tRNA-synt_2c.
CC InterPro; IPR006193; tRNA_synt_Ala.
CC Pfam; PF02272; DHHA1; 1.
CC Pfam; PF01411; tRNA-synt_2c; 1.
```



```

RC STRAIN=cv. Columbia;
RA Desprez T., Anselme J., Chiapello H., Rouze P., Caboche M., Hofte H.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE OF 108-191 FROM N.A.
RC STRAIN=cv. Columbia;
RA Berthomieu P., Guerrier D., Giraudat J.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
[4]
RN
RP SEQUENCE OF 197-407 FROM N.A.
RC STRAIN=cv. Columbia;
RA Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the L4E family of ribosomal proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AL162973; CAB86041.1; -.
CC EMBL; Z18118; CAA79104.1; -.
CC EMBL; Z18460; CAA79191.1; -.
CC EMBL; Z17589; CAA79004.1; -.
CC PIR; T48308; T48308.
CC InterPro; IPR002136; Ribosomal_L4/L1E.
CC Pfam; PF00573; Ribosomal_L4; 1.
CC PROSITE; PS00939; RIBOSOMAL_L1E; 1.
CC
CC Ribosomal protein. 181 G -> V (IN REF. 3).
CC CONFLICT 181 E -> D (IN REF. 4).
CC FT CONFLICT 362
CC FT CONFLICT 407 AA; 44722 MW; 026FC2852B7A2038 CRC64;
CC SEQUENCE 407 AA; 44722 MW; 026FC2852B7A2038 CRC64;
-----
Query Match 7.38; Score 78; DB 1; Length 407;
Best Local Similarity 26.48; Pred. No. 46;
Matches 29; Conservative 24; Mismatches 41; Indels 16; Gaps 4;
QY 1 MKKFLIIAAVFAVVSADPT---HYDKITTEINKAIDDAIAAEQS----ETIDPMKVPD 53
DB 125 MKRHAIVSAIAATAVPALVMARGHKIENFPEPLVWSDSAEAEKTSAAIKVLKQIGAYD 184
QY 54 HADKFERHGVIVDPKGLAMENTEARGLKOMKQGDANVGEEG--IVKA 101
DB 185 DAEKAKNSIGIRPGKGV-----RNRYSIRKGLPLVFTGEGAKIVKA 227
-----
RESULT 91
CH60_RICHS STANDARD; PRT; 555 AA.
ID CH60_RICHS STANDARD; PRT; 555 AA.
AC P16625;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Major antigen 58)
DE (58 kDa antigen).
DE GROL OR GROEL OR MOPA OR STAS8.
OS Rickettsia tsutsugamushi.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Orientia.
OC NCBI_TaxID=784;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216005; PubMed=2108930;
RA Stover C.K., Marxana D.P., Dasch G.A., Oaks E.V.;
RT "Molecular cloning and sequence analysis of the Sta58 major antigen
RT gene of Rickettsia tsutsugamushi: sequence homology and antigenic
RT comparison of Sta58 to the 60-kilodalton family of stress proteins.";
RL Infect. Immun. 58:1360-1368(1990).
CC -1- FUNCTION: Prevents unfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress

```

```
CC conditions (By similarity).
CC -! SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -! SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31887; AAA26393.1; -
DR PIR; B41492; B41492.
DR HSP; P06139; 1GRL.
DR HAMAP; MF_00600; -; 1.
DR InterPro; IPR001844; Chaperonin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PROSITE; PS00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Antigen.
KW SEQUENCE 555 AA; 59729 MW; 16400249D3FC03B4 CRC64;
Query Match 7.3%; Score 78; DB 1; Length 555;
Best Local Similarity 25.3%; Pred. No. 66;
Matches 40; Conservative 19; Mismatches 77; Indels 22; Gaps 4;
CC
CC 18 DPHYDKITEINKAIDATAAIIQSETIDPMKVPDHDKFERHVGIVDFKGLAMNIE 77
CC 401 DALRATRAAEVGVGGVAFVSRVLSLKP-----DNEQDORVGI-----NII 446
CC
CC 78 ARGILKMKRQGDANVKGEEGIVKAHLIGVHD-----DIVSMYDLYAKLGLDLPHTTHVIS 133
CC 447 KKVLAEPRQIVKVGAGKEDVVNLSKSTDKRGFDARTMQVYDMIKAGIVPTKVVRT 506
CC
CC 134 DIQDFVVALSLEISDEGNITWTFEVEQFANVNVHIG 171
CC 507 ALQDAFVSFASVIATSAMIT-----DHEDNNTGNRSGG 540
CC
RESULT 92
Y4FB RHISN
ID Y4FB RHISN STANDARD; PRT; 664 AA.
AC P55440;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 73.7 kDa protein Y4FB.
GN Y4FB.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Perret C.A., Fellay R., Baroch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -! SIMILARITY: None obvious.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000072; AAB91659.1; -
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001604; Endonuclease.
DR Pfam; PF01223; Endonuclease; 1.
DR SMART; SM00477; NUC; 1.
KW Hypothetical protein; Plasmid.
KW SEQUENCE 664 AA; 73731 MW; CFCC041FB73C064F CRC64;
Query Match 7.3%; Score 78; DB 1; Length 664;
Best Local Similarity 22.6%; Pred. No. 82;
Matches 45; Conservative 27; Mismatches 65; Indels 62; Gaps 9;
CC
CC 28 EINKAIDDAI-----AAIEQSETIDPMKVPDHA-----DKFERHVGIVDFKGE 70
CC 2 EISKFADEQLADQIAAGREAVPEVDPIRLIQVHVRARAGAGVDFVFESLIGRID--- 57
CC
CC 71 LAMRNIEARGLKQMKR-----QGDANVKGEEGIVKAHL-----IGVHDDIVSME 115
CC 58 LVEINYLERGLMASRAVCINVPAPVPGDSDWGTGFLIGPRLLLNNHVIDSAEDALKAT 117
CC
CC 116 YDLAYKL---GDLHPTTH-----VIS--DIQDFVVALSLEISDEGNITWTF--- 157
CC 118 VEEGYELDAEGELKETTFRSLSPQDGFVTSRDALDYTVVAIENSEDGTATPISDFGLR 177
CC 158 -----EVQFANVNVH 168
CC 178 LDARTGKTEVGQYATIIGH 196
CC
RESULT 93
POLIS_IDV5
ID POLIS_IDV5 STANDARD; PRT; 1012 AA.
AC P25219;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Structural polypeptide [Contains: Major structural protein VP2;
DE Nonstructural protein VP4; Minor structural protein VP3].
OS Avian infectious bursal disease virus (strain 52/70) (IBDV).
OC Viruses; dsRNA viruses; Birnaviridae; Avibirnavirus.
OX NCBI_TaxID=10996;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90278420; PubMed=2161902;
RA Bayliss C.D., Spies U., Shaw K., Peters R.W., Papageorgiou A.,
RA Mueller H., Boursnell M.E.G.;
RT "A comparison of the sequences of segment A of four infectious bursal
RT disease virus strains and identification of a variable region in
RT VP2.";
RL J. Gen. Virol. 71:1303-1312(1990).
CC -! FUNCTION: Segment A encodes a polypeptide that is processed into
CC the major structural proteins of the virion VP2 and VP3, and into
CC the putative protease VP4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00869; BAA00745.1; -
DR PIR; JQ0941; GNXS52.
DR MEROPS; S50.002; -
DR InterPro; IPR002662; Birna_VP2.
DR InterPro; IPR002663; Birna_VP3.
DR InterPro; IPR002664; Peptidase_S50.
DR Pfam; PF01766; Birna_VP2; 1.
DR Pfam; PF01767; Birna_VP3; 1.
CC
```


Search completed: August 6, 2004, 16:00:53
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 15:57:51 ; Search time 38 Seconds

(without alignments)
1768.563 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068

Sequence: 1 MKKFLIIAAVAFVAVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SPTRMBL25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	109	10.2	568	10 Q9LXF5	Q9LXF5 arabidopsis
2	103	9.6	436	2 Q9RNV5	Q9RNV5 streptococc
3	103	9.6	436	16 Q8DS90	Q8DS90 streptococc
4	100.5	9.4	531	5 Q8I2H1	Q8I2H1 plasmodium
5	99	9.3	260	5 Q9V3A0	Q9V3A0 drosophila
6	97.5	9.1	776	16 Q98PT2	Q98PT2 mycoplasma
7	96.5	9.0	544	16 Q8FNV8	Q8FNV8 corynebacte
8	96	9.0	234	4 Q8N4C7	Q8N4C7 homo sapien
9	96	9.0	959	16 Q8IWA0	Q8IWA0 bacillus an
10	95.5	8.9	186	2 Q8KVC7	Q8KVC7 uncultured
11	95	8.9	436	16 Q8E3T9	Q8E3T9 streptococc
12	93.5	8.8	449	17 Q9HR60	Q9HR60 halobacteri
13	93.5	8.8	628	17 Q9V1N1	Q9V1N1 pyrococcus
14	93	8.7	436	16 Q8DY73	Q8DY73 streptococc
15	92.5	8.7	637	10 Q8XIJ5	Q8XIJ5 arabidopsis
16	92.5	8.7	859	16 Q7VJY3	Q7VJY3 helicobacte

17	92	8.6	186	2 Q8KVM3	Q8KVM3 uncultured
18	92	8.6	186	2 Q8KV02	Q8KV02 uncultured
19	92	8.6	186	2 Q8KHA5	Q8KHA5 uncultured
20	91.5	8.6	186	16 Q98PK8	Q98PK8 mycoplasma
21	91	8.5	186	2 Q8KV56	Q8KV56 uncultured
22	91	8.5	186	2 Q8KVJ7	Q8KVJ7 uncultured
23	91	8.5	444	2 Q84H55	Q84H55 anaplasma m
24	90.5	8.5	186	2 Q8KV52	Q8KV52 uncultured
25	90.5	8.5	681	5 Q8I126	Q8I126 caenorhabdi
26	90	8.4	186	2 Q8KVE0	Q8KVE0 uncultured
27	90	8.4	186	2 Q8KVE1	Q8KVE1 uncultured
28	90	8.4	333	16 Q8DD69	Q8DD69 vibrio vuln
29	90	8.4	1558	5 Q96275	Q96275 plasmodium
30	90	8.4	1786	5 Q9U0P0	Q9U0P0 plasmodium
31	89.5	8.4	186	2 Q8KVL4	Q8KVL4 uncultured
32	89.5	8.4	939	2 Q8KMS5	Q8KMS5 marionomonas
33	88.5	8.3	279	16 Q7MNV1	Q7MNV1 bordetella
34	88.5	8.3	279	16 Q7W160	Q7W160 bordetella
35	88.5	8.3	1393	16 Q822J2	Q822J2 chlamydomophi
36	88.5	8.3	2137	10 Q9SVN7	Q9SVN7 arabidopsis
37	88	8.2	609	16 Q8FGF6	Q8FGF6 escherichia
38	87.5	8.2	485	16 Q897K6	Q897K6 clostridium
39	87.5	8.2	1620	5 Q8IK83	Q8IK83 plasmodium
40	87	8.1	186	2 Q8KV75	Q8KV75 uncultured
41	87	8.1	372	10 Q9ZVB0	Q9ZVB0 arabidopsis
42	87	8.1	385	10 Q8GWB1	Q8GWB1 arabidopsis
43	87	8.1	399	5 Q44999	Q44999 caenorhabdi
44	87	8.1	562	17 Q8U2M6	Q8U2M6 pyrococcus
45	87	8.1	1640	9 Q9AZS0	Q9AZS0 bacterioph
46	87	8.1	1640	16 Q9CF57	Q9CF57 lactococcus
47	86	8.1	280	16 Q7TZ25	Q7TZ25 mycobacteri
48	86	8.1	303	16 Q9I432	Q9I432 pseudomonas
49	86	8.1	339	16 P71763	P71763 mycobacteri
50	86	8.1	1661	5 Q96166	Q96166 plasmodium
51	85.5	8.0	186	2 Q8KV93	Q8KV93 uncultured
52	85.5	8.0	186	2 Q8KVA8	Q8KVA8 uncultured
53	85.5	8.0	263	4 Q86YC9	Q86YC9 homo sapien
54	85.5	8.0	264	9 Q7Y444	Q7Y444 enterobacte
55	85.5	8.0	461	16 Q7VFQ7	Q7VFQ7 helicobacte
56	85.5	8.0	745	4 Q8N3E4	Q8N3E4 homo sapien
57	85.5	8.0	807	16 Q7UYX3	Q7UYX3 rhodopirell
58	85.5	8.0	990	16 Q8EAY1	Q8EAY1 shewanella
59	85.5	8.0	1062	4 Q7Z619	Q7Z619 homo sapien
60	85.5	8.0	1154	11 Q91ZS0	Q91ZS0 mus musculu
61	85.5	8.0	1441	4 Q60310	Q60310 homo sapien
62	85.5	8.0	2238	11 Q91ZS1	Q91ZS1 mus musculu
63	85.5	8.0	2780	11 Q91ZS2	Q91ZS2 mus musculu
64	85.5	8.0	6298	11 Q8VHN7	Q8VHN7 mus musculu
65	85	8.0	185	-2 Q8KVK6	Q8KVK6 uncultured
66	85	8.0	186	2 Q8KVC0	Q8KVC0 uncultured
67	85	8.0	186	2 Q8KV82	Q8KV82 uncultured
68	85	8.0	292	17 Q9UZS6	Q9UZS6 pyrococcus
69	85	8.0	401	16 Q8RDZ2	Q8RDZ2 fusobacteri
70	85	8.0	537	16 Q8DBL1	Q8DBL1 vibrio vuln
71	85	8.0	570	16 Q8F017	Q8F017 leptospira
72	85	8.0	878	16 Q9CFU9	Q9CFU9 lactococcus
73	85	8.0	334	2 Q8RQU4	Q8RQU4 thermus the
74	84.5	7.9	426	16 Q8AOK9	Q8AOK9 bacteroides
75	84.5	7.9	710	16 Q8PG51	Q8PG51 xanthomonas
76	84.5	7.9	817	16 Q8CQ88	Q8CQ88 staphylococ
77	84.5	7.9	841	16 Q898W1	Q898W1 clostridium
78	84.5	7.9	1429	5 Q8SXM3	Q8SXM3 drosophila
79	84.5	7.9	186	2 Q8KV19	Q8KV19 uncultured
80	84	7.9	186	2 Q8KV15	Q8KV15 uncultured
81	84	7.9	306	16 Q8DWU5	Q8DWU5 streptococc
82	84	7.9	314	16 Q9WZQ7	Q9WZQ7 thermotoga
83	84	7.9	392	16 Q99X76	Q99X76 staphylococ
84	84	7.9	600	2 Q9X916	Q9X916 versinia en
85	84	7.9	600	16 Q92375	Q92375 versinia pe
86	84	7.9	775	16 Q9CHZ3	Q9CHZ3 lactococcus
87	84	7.9	9510	2 Q93NX9	Q93NX9 streptomyce
88	84	7.9	186	2 Q8KVH7	Q8KVH7 uncultured
89	83.5	7.8			

90 83.5 7.8 266 16 Q7VTQ1
91 83.5 7.8 311 2 Q93R43
92 83.5 7.8 311 2 Q93R44
93 83.5 7.8 311 2 Q93R42
94 83.5 7.8 329 2 Q9EXN7
95 83.5 7.8 402 11 Q8K3U8
96 83.5 7.8 412 16 Q8NUB6
97 83.5 7.8 474 16 Q97K37
98 83.5 7.8 1047 4 Q9C0B4
99 83.5 7.8 1302 4 Q7Z2L3
100 83 7.8 186 2 Q8KVP2

ALIGNMENTS

RESULT 1
Q9LXF5 PRELIMINARY; PRT; 568 AA.
AC Q9LXF5
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Hypothetical protein.
GN F8W21_160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AL353993; CAB89337.1; -.
DR PIR; T49962; T49962.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH; 4.
DR SMART; SMO0322; KH; 4.
DR PROSITE; PS00084; KH_TYPE_1; 4.
KW Hypothetical protein.
SQ SEQUENCE 568 AA; 62034 MW; 3431FB92B4E37205 CRC64;

Query Match 10.2%; Score 109; DB 10; Length 568;
Best Local Similarity 26.7%; Pred. No. 1.5;
Matches 46; Conservative 36; Mismatches 64; Indels 26; Gaps 8;
QY 25 ITEINKAIDDAIAAEQSEITDPMK--VPDHAKFERHVGIVDFKGLAMRN-----IE 77
DB 331 ITISAREVFEDA----YSPTEAVMRLOPKCSKVERDSGLVSFTTLLVPSRIGCIL 385
QY 78 ARG---LQMKRGQDANVK--GEGIVKA-----HLIGVHDDIVSMYDLYAYKGLDHP 128
DB 386 GKGGAIITETRMRTKANIRILGKLENLKVASDDDDMMQVNFVWMLLLKFSLOFLLSRLRYF 445
QY 129 THVTSIDIDFVVALLS--LEISDEGNITWTSFEVRFQFANVNHIGGLSILDPI 178
DB 446 TWSVNSNSNFQVQISGELDVAKELIQITS---RLRNVFDEGAVSALAPV 494

RESULT 2
Q9RHV5 PRELIMINARY; PRT; 436 AA.
AC Q9RHV5
DT 01-MAY-2000 (TREMREL. 13, Created)

DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-OCT-2002 (TREMREL. 22, Last annotation update)
DE Phosphoglycerate dehydrogenase.
GN PGDA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT8148;
RA Kawabata S., Terao Y., Hamada S.;
RT "Molecular cloning, sequence and characterization of a novel
RT streptococcal phosphoglycerate dehydrogenase gene.";
RL Oral Microbiol. Immunol. 15:58-62 (2001).
DR EMBL; AB016077; BAA88823.1; -.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR006073; GTP_OBG.
DR InterPro; IPR002917; MMR_HSR1.
DR Pfam; PF01926; MMR_HSR1; 1.
DR PRINTS; PR00326; GTP_OBG.
DR TIGRFAMs; TIGR00650; MG442; 2.
DR TIGRFAMs; TIGR00231; small_GTP; 2.
SQ SEQUENCE 436 AA; 48601 MW; EC52CC8650DED090 CRC64;
Query Match 9.6%; Score 103; DB 2; Length 436;
Best Local Similarity 23.0%; Pred. No. 3.2;
Matches 50; Conservative 36; Mismatches 77; Indels 54; Gaps 11;
QY 30 NKAIDDAIAAEQSEITDPMKVPDHAKFERHVGIVDFKGLAMRNIEARGLKQMKRQGD 89
DB 21 NRIAGERISIVDEVEGVTRDRIYTKAEWLNRQFSIIDTGG---IDVDAPFMEQIKHQAD 77
QY 90 AN-----VKGEIGIVK-----AHLIGVHDDIV-----SMEYDL-AYK 121
DB 78 IATMEADVIVFVSAKEGITDADEVAKILYRTHKPVILAVKNVDNPEMSAIYDFVALG 137
QY 122 LGDLHP--TTHVI--SDIQDFTVW-ALSLEISDEGNITWTSFEVRF-----ANVY 166
DB 138 LGDPFVSSAHGIGTGVDLDAIVNLPAAQEE-----SSDIKFSLIGPNVCKSSLI 191
QY 167 NHIGLS--ILDPIFGLSDVLTAFQDTVRKEMTKV 201
DB 192 NAILGEDRVIASPVAGTTRDAITDTFTDEGEQFTMI 228
RESULT 3
Q9DS90 PRELIMINARY; PRT; 436 AA.
AC Q9DS90
DT 01-MAR-2003 (TREMREL. 23, Created)
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Phosphoglycerate dehydrogenase.
GN PGDA OR SMU.1920.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AF015016; AAN59531.1; -.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR006073; GTP_OBG.

DR InterPro; IPR002917; MMR_HSR1.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PFO1926; MMR_HSR1; 1.
DR PRINTS; PRO0326; GTF1OBG.
DR TIGRFAMs; TIGR00650; MG442; 2.
DR TIGRFAMs; TIGR00231; small_GTP; 2.
KW Complete proteome.
SQ SEQUENCE 436 AA; 48585 MW; 24D8428A91C2A097 CRC64;

Query Match
Best Local Similarity 9.6%; Score 103; DB 16; Length 436;
Matches 50; Conservative 36; Mismatches 77; Indels 54; Gaps 11;

QY 30 NKAIDDDAAIALEQSETDPMKVPDHADKFERHVGIVDFKGLAWRNTEARGLKOMKRGD 89
DB :
21 NRAGEISIVEDVEGTORIRIYTKAELNKFQSIIDTGG---IDDVDAPPMEIQKHQAD 77
QY 90 AN-----VKGEEGIVK-----AHLITGVGDIV-----SMEDYL-AYK 121
DB :
78 IAMTEADVIVFVVSAGKIGTDADDEVAKILYTHKPVILAVKNVDNPEMSAIYDFVALG 137
QY 122 LGDLHP--THVI--SDIQFVV-ALSLEISDEGNITWTSFEVRQF-----ANVV 166
DB :
138 LGGDPYPVSSAHGIGTGVDLDAIVNLPTAEQEE-----SSDIIFKFLGRPNVNGKSLL 191
QY 167 NHIGGLS--ILDPIFGVLSDVLTAIFOOTVRKEMTKV 201
DB :
192 NAIIGEDRVIASPVAGITRDIAIDTFTDEEQEFTMI 228

RESULT 4
Q812H1 PRELIMINARY; PRT; 531 AA.

AC Q812H1; 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF1680W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=36329;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Falwell T., Gobie A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch Z.,
RA Rajandream V.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929359; CAD52022.1; -.
DR InterPro; IPR006577; UAS.
DR InterPro; IPR001012; UBX.
DR Pfam; PFO0789; UBX; 1.
DR SMART; SM00594; UAS; 1.
DR PROSITE; PS50033; UBX; 1.
KW Hypothetical protein.
SQ SEQUENCE 531 AA; 62569 MW; 7D1867D7919C7A11 CRC64;

Query Match
Best Local Similarity 9.4%; Score 100.5; DB 5; Length 531;
Pred. No. 6.8;

```

RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=Berkley;
EX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazer R., Davis T.,
RA Doyle C., Galle R., Harris N., Hartzell G., Harvey D.,
RA Horg L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Teang G., Wan K., Whitelaw K.,
RA Celniker S., Rubin G.M.;
RA "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazer R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.R., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lonoten M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RL EMBL; AE003641; AAF53345.1; -
DR EMBL; AE003408; AAF44828.1; -
DR Flybase; FBgn0028920; Bg:DS00941.12.
DR InterPro; IPR004272; Odorant_binding.
DR Pfam; PF03027; DUF233; 1.
DR SMART; SM00700; JHBP; 1.
KW Hypothetical protein.
SQ SEQUENCE 260 AA; 28095 MW; FC3486D66A29A6AB CRC64;

Query Match 9.3%; Score 99; DB 5; Length 260;
Best Local Similarity 21.9%; Pred. No. 3.5;
Matches 56; Conservative 41; Mismatches 63; Indels 96; Gaps 13;

QY 5 LLIAAVAFVAVSA-----DPHYDKITEINKAIDDAIAIEQS-----ETIDPM 49
DB 3 IFVAILAFVAVSAASNGQPIET---TQSSITVDVIEGKEQMGCGTSVGLPLAPL 58
QY 50 KVPDHAKFERHVGIVDPKGLAVNTEARGLKQKQGDANVKGEGIVKAHLIGVHD 109
DB 59 RI-DHQD-----INI-----DSSVLKAGQGTIDHFLNGLND 88
QY 110 -DI-----VSMGYDLAYKGLDGLHPTTHVIS-----DIQDFV 140
DB 89 FDIEMKNALTSKTYKFTEDVNVDTQDLSVLLKKYGTINLIGAGHAKFAIKDWI 148
QY 141 ALSLEIS---DEGNTMTSFEVRFQFANVNVHIGLSILDPIFGVLSVLFAIFQDTVRKE 197
DB 149 WGTWKYSLGVISGNLKLKSLVR-----THLG--EVDSEIEGILGD-----GSINEK 193
QY 198 MTKVLAPAFKRELEKN 213
DB 194 MNEYLAELAEVELAINEN 209

RESULT 6
Q98PT2 PRELIMINARY; PRT; 776 AA.
AC Q98PT2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE LIPOPROTEIN.
GN MYPU 6370.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]

SEQUENCE FROM N.A.
STRAIN=UAB CTIP;
MEDLINE=21267165; PubMed=11353084;
Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
Blanchard A.;
"The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445565; CAC13810.1; -
DR PIR; S90591; E90591.
DR MyPulist; MYPU 6370; -
DR InterPro; IPR00437; Prok_lipoprot_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 776 AA; 88710 MW; 94F1B4A485FA4DB4 CRC64;

Query Match 9.1%; Score 97.5; DB 16; Length 776;
Best Local Similarity 20.1%; Pred. No. 20;
Matches 52; Conservative 37; Mismatches 73; Indels 97; Gaps 10;

QY 22 YDKITEINKAIDDAIA---AIEQSETIDPMKVPD-HADKFE-----RHVGIVD--- 66
DB 437 YNQRFDQVNLKLEQAFFPKNSWEATQNPSPLDENSTYQDKLDNIGDEKAKFIGITDLP 496
QY 67 -----FKGELAMRN-----IEARGLKQKQGDANVKGEGIVK-----A 101
DB 497 NELKSLEFKKATALRNATAIIGLNTLAELKKGESQASRDILVKADFEKEMQHDNDNA 556
QY 102 HL-----LIGVHDDIVSMGYDLAYKGLDGLHPTTHVISDIQDFVVALSLEIS----- 147
DB 557 HURKTEKIFGVNGDEQSDKDDLVKQLSKQSFISAVESIH---AKALESRKYAKNV 612
QY 148 DEGNI-----TMTSFEVRFQFANVNVH 168
DB 613 DENNIKAVELYLANYNQAIWNKFKNKYLEIKTWTSKFKKDKGQSDSSMTIEQVINEYK 672
QY 169 IGLSILDPFGVLSVL 187
DB 673 YRSSNVADKIISVITSTNT 691

RESULT 7
Q98PT2 PRELIMINARY; PRT; 544 AA.
AC Q98PT2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter ATP-binding protein.
GN CE2035.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
SEQUENCE FROM N.A.
STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
Usuda Y., Sugimoto S.;
"The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF052221; BAC18845.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
RN [1]

```

DR PROSITE; PS00211; ABC TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Complete; Proteome
SQ SEQUENCE 544 AA; 59593 MW; 3BE49CA60450A53E CRC64;

Query Match
Best Local Similarity 9.0%; Score 96.5; DB 16; Length 544;
Matches 47; Conservative 30; Mismatches 64; Indels 51; Gaps 10;

QY 19 PIHYDKITEE-----INKAIDDAIAAEQSETI-----DPMKVPDHDKPERH 61
Db 67 PETHGTAQTDLPFDLPVQSLIDEAVPRAVETTIGELTVAMGDDNPPDIAERFDBA 126
QY 62 VGVDFKGLAMENIEAR-----GLKMKQGDANVKGBEG-----IVKAHLILIGVH 108
Db 127 LSRAE---ELGVNLEARIEIEI VAGLGLADVDK--SVPIRALSGGQRRRFALAVILLEPH 181
QY 109 DDIVSMEYDLAYKGLDHPHTHVISDIQDFVALSLEISDEGNITMTSEVRFQFANVNH 168
Db 182 DAILDE-----PTNHLDDNAVDFLI--NELTSFKGPVLITASHD--RFFLDVAV-- 225
QY 169 IGGSLIDPIFG 180
Db 226 VTGLVDLPSLG 237

RESULT 8
Q8N4C7 PRELIMINARY; PRT; 294 AA.
ID Q8N4C7
AC Q8N4C7
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Similar to RIKEN CDNA A030009B12 gene.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034696; AAH34696.1; -
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0008565; P:protein transporter activity; IEA.
DR GO; GO:0006886; P:intracellular protein transport; IEA.
DR InterPro; IPR006012; Syntaxin.
DR InterPro; IPR000727; T SNARE.
DR InterPro; IPR006011; Syntaxin N.
DR Pfam; PF05739; SNARE; I.
DR Pfam; PF08004; Syntaxin; 1.
DR SMART; SM00397; t-SNARE; 1.
DR SMART; SM00914; Syntaxin; 1.
DR PROSITE; PS00914; SYNTAXIN; 1.
DR PROSITE; PS0192; T SNARE; 1.
SQ SEQUENCE 294 AA; 34324 MW; 1A7E986916DCE57 CRC64;

Query Match
Best Local Similarity 9.0%; Score 96; DB 4; Length 294;
Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;

QY 27 EEINKAIDDAIAAEQSETI-----PDHAKFERHVGIVDF-----KGE- 70
Db 103 EYINRSLNDLVKKEKSEVENGPPSSVVTILKSQAAMP-RHFQIMFYNDTIAAKQEK 161
QY 71 ---LAVNTEARG-----LKMKEQGDANVKGBEG-----GLVKAHL---LGVHDDIVS 113
Db 162 CKTFILRQLEAVAKEMSEEDVNDMLHQKWEVFNESLLTEINITKAQSEISQRKELVN 221
QY 114 MEYDLAYKGLDHPHTHVISDIQDFVALSLEISDEG-----NITMTSEVRFQFANVNH 169
Db 222 LE-----NQIKLRDLFIQISLLVEEGESINNIENTVNSTKEVYVNTKEK 267

QY 170 GGLST 174
Db 268 FGLAV 272

RESULT 9
Q81WA0 PRELIMINARY; PRT; 959 AA.
ID Q81WA0
AC Q81WA0
DT 01-JUN-2003 (TREMblrel. 24, Created)
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)
DE Prophage Lambda02, tape measure protein, putative.
GN BA4082.
OS Bacillus anthracis (strain Ames)
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rillstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy K.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman J.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
DR EMBL; AE017036; AAP27807.1; -
DR TIGR; BA4082; -
KW Complete proteome.
SQ SEQUENCE 959 AA; 104502 MW; F70773DD6A500055 CRC64;

Query Match
Best Local Similarity 9.0%; Score 96; DB 16; Length 959;
Matches 51; Conservative 33; Mismatches 57; Indels 76; Gaps 11;

QY 23 DKITEINKAIDDAIAAEQSETI-----DPMKVP-DHAKFERHVGIVDFKGE- 70
Db 764 NVVIESINKEIEVA-----TQAEKLEEGIKRYQDGLNPDFAQKFEQALQVADRNVQ 819
QY 71 -----LAMRNIEARGKQKQGDANVKGBEGIVKAHLIGVHDDIVSME----- 115
Db 820 FYTKAKETISISKEIEAGGMSL-----DAGKRFESIIKVY-----EDGVKSLDKQTKGW 870
QY 116 ---YDLAYKGLDHPHTHVISDIQDFVALSLEISDEGNITMTSEVRFQFANVNHIGGL 172
Db 871 RENYERAFKLGTEIKP-----EERKATLDAIALYE-----SKHVNDL 906
QY 173 -SILDPIFGVLSVDLTAIFQDITVRKMTKVLAPAFKR 208
Db 907 QSIENDGFKVL-----QOHMKEDAEVLAQAKR 935

RESULT 10
Q8KVC7 PRELIMINARY; PRT; 186 AA.
ID Q8KVC7
AC Q8KVC7
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE 60 kDa chaperonin (fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]

RESULT 13
Q9V1N1


```

DR GO: 0005524; F:ATP binding; IEA.
DR GO: 0003754; F:Chaperone activity; IEA.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR InterPro: IPR008950; GroEL-ATPase.
DR Pfam: PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 186
FT NON_TER 186
SQ SEQUENCE 186 AA; 20216 MW; D08674022E4E6732 CRC64;

Query Match      8.6%; Score 92; DB 2; Length 186;
Best Local Similarity 25.4%; Pred. No. 8.6;
Matches 30; Conservative 23; Mismatches 31; Indels 34; Gaps 6;

QY 13 VAVSADPIHYDKITEINKAIDAIAAI-----EQSETIDPMKVPDPHADKFER--HVGIV 65
DB 16 VTAGANPM-----DLKRGIDKAVAKVVENIAEQAEV-----GDKFEKIEHVAKI 60

QY 66 DFKGELAMRNIEARGLKQKMRQGDANV---KGEGIVKAHLIGVHDDIVSMEYDLAY 120
DB 61 SANGDEAIGKLIAMQVRKKGVIIVBEAKGTETTV-----DVEGQFDRGY 109

RESULT 19
Q8KHA5 PRELIMINARY; PRT; 186 AA.
AC Q8KHA5;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN Cpn60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
  Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
  throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR ENBL; AF436928; AAM49189.1; -
DR ENBL; AF436929; AAM49190.1; -
DR ENBL; AF436936; AAM49197.1; -
DR ENBL; AF436963; AAM49224.1; -
DR ENBL; AF436980; AAM49241.1; -
DR ENBL; AF437078; AAM49339.1; -
DR ENBL; AF437182; AAM49443.1; -
DR ENBL; AF437184; AAM49445.1; -
DR ENBL; AF437285; AAM49546.1; -
DR GO: 0005524; F:ATP binding; IEA.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR InterPro: IPR008950; GroEL-ATPase.
DR Pfam: PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 186
FT NON_TER 186
SQ SEQUENCE 186 AA; 20167 MW; D0867410A54E6732 CRC64;

Query Match      8.6%; Score 92; DB 2; Length 186;
Best Local Similarity 25.4%; Pred. No. 8.6;
Matches 30; Conservative 23; Mismatches 31; Indels 34; Gaps 6;

QY 13 VAVSADPIHYDKITEINKAIDAIAAI-----EQSETIDPMKVPDPHADKFER--HVGIV 65
DB 16 VTAGANPM-----DLKRGIDKAVAKVVENIAEQAEV-----GDKFEKIEHVAKI 60

QY 66 DFKGELAMRNIEARGLKQKMRQGDANV---KGEGIVKAHLIGVHDDIVSMEYDLAY 120
DB 61 SANGDEAIGKLIAMQVRKKGVIIVBEAKGTETTV-----DVEGQFDRGY 109

```

```

RESULT 20
Q98PK8 PRELIMINARY; PRT; 979 AA.
ID Q98PK8;
AC Q98PK8;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE P115-like (Mycoplasma hyorhinis) ABC transporter ATP-binding
  protein.
DE protein.
GN MYPU_7140.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chamberaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
  Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
  Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
  Mycoplasma pulmonis";
RL Nucleic Acids Res. 29:2145-2153 (2001).
DR ENBL; AL445565; CACI3887.1; -
DR PIR; B90601; B90601.
DR MyPUList; MYPU_7140; -
DR GO: 00016020; C:membrane; IEA.
DR GO: 0005634; C:nucleus; IEA.
DR GO: 0005524; F:ATP binding; IEA.
DR GO: 0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO: 0007059; F:chromosome segregation; IEA.
DR GO: 0006810; P:transport; IEA.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR003395; SMC_N.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
KW Complete proteome.
SQ SEQUENCE 979 AA; 110333 MW; 4BC5C4BE1EA54ACE CRC64;

Query Match      8.6%; Score 91.5; DB 16; Length 979;
Best Local Similarity 21.2%; Pred. No. 84;
Matches 46; Conservative 45; Mismatches 93; Indels 33; Gaps 9;

QY 3 KFLIIAFAVAVSADPIHYDKIT-----EINKAIDAIAIAEQSETIDPMK 50
DB 616 EFSVEAAKVEKLSNDQKSFDSLTAETQTISAKDLGESSSSALDQQINEIQAKKVI--LQ 673

QY 51 VPDHADKFERHVGIVDFKGLAMENIEARGL-KMKRQGDANVKGEGIVKAHLIGVHD 109
DB 674 AQIRAK-----ESILSLNTDFSKVTLEKNLLEKSLRLINDSFKKITEKNAEFYLDNQ 729

QY 110 DIVSMEYDLAYKLG-DLHPTTHVISDIQDFYVALSLSEISDEGNITMTSFVRQFANYVNH 168
DB 730 KRLSEFYNTYEMAKDNFSLIDITEAEAEFVGRKEIKDLGNINIDS--IKSYEISER 787

QY 169 IGGILDPFEGVLSVLTAFQDVTVR---KENTKVL 202
DB 788 YDGLKATE-----SEIVDA--KDKIESATKEMDKII 816

RESULT 21
Q8KV56 PRELIMINARY; PRT; 186 AA.
ID Q8KV56;
AC Q8KV56;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN Cpn60.

```



```

DR InterPro: IPR008950; GroEL-ATPase.
DR Pfam: PF00118; cpm60_TCP1; 1.
DR ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20156 MW; A114C4B4FAB8600 CRC64;

Query Match      8.5%; Score 90.5; DB 2; Length 186;
Best Local Similarity 24.2%; Pred. No. 11;
Matches 30; Conservative 23; Mismatches 44; Indels 27; Gaps 5

QY 7 IAAVAFVAVSADPIHYDKITEINKAIDDAIAAI-----EQSETIDPMKVPDHADKFER- 60
DB 3 VLAQAIIGVLGNVTADANPMDLKGIDKAVAKVVENIAEQAEV-----GDKFEKI 54
QY 61 -HVGIVDPFGELAMENIEARGLKQWKQGGDANV---KGEIGIVKAHLITGVHDDIVSMY 116
DB 55 EHVAKISANGDETIGKLIARAWGKVKKEGVIIVTEEAKGTETTV-----DWVEGMOF 105
QY 117 DLAY 120
DB 106 DRGY 109

RESULT 25
Q18126 PRELIMINARY; PRT; 681 AA.
ID Q18126 AC Q18126;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN C24A3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1] RN
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2] RN
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favell A.;
RT "The sequence of C. elegans cosmid C24A3.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3] RN
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U40424; AAA81461.1; -.
DR PIR; T15590; T15590.
DR WormPep; C24A3.1; CE04058.
DR InterPro: IPR008627; GETHR.
DR Pfam; PF05671; GETHR; 1.
DR Hypothetical protein.
SQ SEQUENCE 681 AA; 79533 MW; 2DA37C9046C26A20 CRC64;

Query Match      8.5%; Score 90.5; DB 5; Length 681;
Best Local Similarity 25.1%; Pred. No. 63;
Matches 51; Conservative 41; Mismatches 62; Indels 49; Gaps 11

QY 24 KITEINKAID-----AIAAIE-----QSETI-----DPMKVPDHADKFERHVGIVDFKGL 71
DB 133 KNTENQQAQKDFNSRIGVLSRMSIRASERIVANENMR-----HSFEAKLAELSQLEL 187

```

Qy	72	AVRNEIARGLKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMYDIAVKLGDLHP	THV 131
Dy	188	AVKQITSEKAKQKQK-----FQKNEALAALEHHL--ELGNSKIDKLM	228
Qy	132	ISDIODFVVALSLETSDEGNIT-MTSFEVROFANVNVNHIGGLS-ILDP	IFGVLSDV-LTA 188
Dy	229	NSEIQ-----ARKHENGKLAQMTDIEDR-----VNVYGGNNKSIDEMKNGKNVHMPA	278
Qy	189	IFQDTRVRKEMTKVLAPAFKRELE 211	
Dy	279	LOTDALRREBAIAADKNKLSME 301	
RESULT 26			
Q8KVE0			
ID	Q8KVE0	PRELIMINARY;	PRT; 186 AA.
AC	Q8KVE0		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	60 kDa chaperonin (fragment).		
GN	CPN60.		
OS	uncultured pig faeces bacterium.		
OC	Bacteria; environmental samples.		
NCBI_TaxID=190966;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22034982; PubMed=12039767;		
RA	Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,		
RA	Crosby W.L., Hemmingsen S.M.;		
RT	"Extensive profiling of a complex microbial community by high-		
RT	throughput sequencing.";		
RL	Appl. Environ. Microbiol. 68:3055-3066(2002).		
DR	EMBL; AF437066; AAA49327.1;--;		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0003754; F:chaperone activity; IEA.		
DR	InterPro; IPR002423; CPN60/TCF-1.		
DR	InterPro; IPR008950; GroEL-ATPase.		
KW	Pfam; PF00118; cpn60_TCF1; 1.		
KW	ATP-binding; Chaperone.		
FT	NON_TER		
FT	NON_TER 1 186		
SQ	SEQUENCE 186 AA; 20229 MW; DB666F10A4FE7C82 CRC64;		
Query Match 8.4%; Score 90; DB 2; Length 186;			
Best Local Similarity 25.4%; Pred. No. 13;			
Matches 30; Conservative .23; Mismatches 31; Indels 34; Gaps 6			
Qy	13	VAVSADPHYDKITEINKAIDDAIAI-----EQSETIDPMKVPDPAKPER--HVGIV	65
Dy	16	VTAGANPM-----DLKRGIDKAVAKWENIAEQTEV-----GDKFEKIEHVAKI	60
Qy	66	DFKGELAMENIARGLKQMKRQGDANV---KGEGIVKAHLIGVHDDIVSMYDIAV	120
Dy	61	SANGDEATGKLIAMAQKVKEGVITTEAKGTETTV-----DVVEGKQFDRGY	109
RESULT 27			
Q8KVE1			
ID	Q8KVE1	PRELIMINARY;	PRT; 186 AA.
AC	Q8KVE1		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	60 kDa chaperonin (fragment).		
GN	CPN60.		
OS	uncultured pig faeces bacterium.		
OC	Bacteria; environmental samples.		
NCBI_TaxID=190966;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22034982; PubMed=12039767;		
RA	Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,		

RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437065; AAM49326.1; -.

DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.

DR InterPro; IPR002423; Cn60/TCP-1.
DR InterPro; IPR008950; Gc6EL-ATPase.

DR Pfam; PF00118; Cn60_TCP1; 1.
DR ATP-binding; Chaperone.

KW NON_TER 1 1
FT NON_TER 1 186
SQ SEQUENCE 186 AA; 20095 MW; 510847D1395090A CRC64;

Query Match 8.4%; Score 90; DB 2; Length 186;
Best Local Similarity 25.4%; Pred. No. 13;
Matches 30; Conservative 23; Mismatches 31; Indels 34; Gaps 6;

QY 13 VAVSADTHYDKITEEINKAIDDAIAI-----EQSETIDPMKVPDHADKFER--HVGIV 65
DB 16 VTAGANPM-----DLKRGIDKAVAKVENIAQAEEV-----GDKFEXIEHVAKI 60

QY 66 DFKELANRNTAARGLKMKGQGANV---KGEIGVKAHLIGVDDIVSMYDLY 120
DB 61 SANGDEAIGKLIAGAMQVKEGVTVEEAKGTETV-----DVVEGMQFDRGY 109

RESULT 28
QSD59 ID Q8DD69 PRELIMINARY; PRT; 333 AA.

AC Q8DD69;
DT 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Glyceraldehyde-3-phosphate dehydrogenase.

GN V11141.

OS Vibrio vulnificus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrrio.

OC NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RT Chey H.E.;
RL "Complete genome sequence of Vibrio vulnificus CMCP6.";
DR EMBL; AE016800; AAO09615.1; -.

DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.

DR GO; GO:0006096; P:glycolysis; IEA.

DR InterPro; IPR006424; GAPDH-I.

DR InterPro; IPR000173; GAP_dhhydrogenase.

DR Pfam; PF00044; gpdh; 1.

DR Pfam; PF02800; gpdh_C; 1.

DR TIGRfams; TIGR01534; GAPDH-I; 1.

DR PROSITE; PS00071; GAPDH; 1.

KW Complete proteome.

SQ SEQUENCE 333 AA; 36634 MW; 87E1D9698395F681 CRC64;

Query Match 8.4%; Score 90; DB 16; Length 333;
Best Local Similarity 24.2%; Pred. No. 27;
Matches 48; Conservative 35; Mismatches 63; Indels 52; Gaps 9;

QY 59 ERHVGIVDFKG-----ELAMENIARGLKMKRG-----DANKGEGIVKAHLIGV 107
DB 78 QSDASVDWSGCDVVIATGVHRSRFLNQVLEGGVVRVSPVK--EEGI--ANIVGV 134

QY 108 HDDIVSMYD-----LAYKLG-----DLHPTTHVISDQDFVVA---- 141
DB 135 NDEIFNPQHRIVTAASCTTNCIAPVVKVIHEKLGIEQASFTTHDUTNTQILDAPHKD 194

QY 142 -----LSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFGV-LSVLTAFQD 192
DB 195 LRARACGSLIPTTTGSATAI--VEIFPELKORINGHAVRVPLANASLTDIFDVKRD 251

QY 193 TVRKEMTKVLAPAFKREL 210
DB 252 TTAEEVNALKEASOGEL 269

RESULT 29
O96275 ID O96275 PRELIMINARY; PRT; 1558 AA.

AC O96275;
DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE RESA-H3 antigen.

GN PF0015W.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OC NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.";
RT Science 282:1126-1132(1998).

DR EMBL; AE001424; AAC71972.1; -.

DR PIR; B71603; B71603.

SQ SEQUENCE 1558 AA; 175658 MW; E2FC68E9036C81BC CRC64;

Query Match 8.4%; Score 90; DB 5; Length 1558;
Best Local Similarity 23.3%; Pred. No. 2.1e+02;
Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;

QY 22 YDKITEEINKAIDDAIAAEQSEITDPMKVPDHPADKFERHVGIVDFKGLAMRNIARGL 81
DB 747 FNTVLDKVEETVEISGESLENNE---MDKAPFFSEIFDNVKGIGENLITGMFRSITSIV 802

QY 82 KMKQEGDANVKGEGIVKAHLIGVDDIVSMYDLYKGLDHPHPTTHVISDQDFVVA 141
DB 803 IOSEKVDLN---ENVVSSIL---DNIEKKEGLLNKLENISSTEGVQETVTEHV-- 851

QY 142 LSLEISDEGNITMTSFEV---RQFANVNVHIGLS---ILDPFGVLSVLTAF--IFQ 191
DB 852 -----EQNV-YVDVDVPAMKQDFLGILNEAGGLKEMFFNLEDVFKSESVDITVEBIKD 903

QY 192 DTVRKEMTKVLAPAFKRELEKN 213
DB 904 EPVQKEVEKETVSIIE-EMEN 924

RESULT 30
Q9U0P0 ID Q9U0P0 PRELIMINARY; PRT; 1786 AA.

AC Q9U0P0;
DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Liver stage antigen-3 precursor.

GN LSA-3.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OC NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=KL;
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;


```

QY 97 GIVKAHL-LIGVHDDIVSMEYDLAYKLGDLHPTT-----HVTSIDQDF-----VVA 141
D 303 --IRTPLNAINGY---IQLCKQDLKQORLYMDTIAQSTTNLITLITSDILDSKIEAGK 357
QY 142 LSLRISDEG-NITMTSFEVROFQVNVHIGGSLDPIFGVLSVLTAFQDTV--RKE 197
D 358 LSLDLTDINIRDI VDEVIQILSSANLVNASKELDLI-PEFDL--DVTWTWIGDPVIRQV 414
QY 198 MTKVLAPAFK 207
D 415 LTNLIGNAIK 424

RESULT 33
Q7WNV1 PRELIMINARY; PRT; 279 AA.
AC Q7WNV1;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE Putative exported protein.
GN BB0934.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-598;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640439; CAE31433.1; -.
KW Complete proteome.
SQ SEQUENCE 279 AA; 30292 MW; 0092B34E2CD15E26 CRC64;

Query Match 8.3%; Score 88.5; DB 16; Length 279;
Best Local Similarity 24.4%; Pred. No. 28;
Matches 53; Conservative 33; Mismatches 74; Indels 57; Gaps 12;

QY 3 KFLIIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-ATAAIEQSETID---PMK 50
D 30 RYLATAVVALAVAAALYARQDPEHFDEV---VLQCLDQAMSTVRRGSDGILPAC 86
QY 51 VPDHADFEHVGIVDFKGLAMENIAR-GLKQMKQGDANVKGEGIVKAHL-----104
D 87 QPELAELVGEDFGAVD---EPTLRVLATVAAANFAEYSSAVAYEAIEKSOHLNCGNL 143
QY 105 -----IGVHDDIVSMEYDLAYKLGD-----LHPTTHVTSIDQDFVAL 142
D 144 VFLMGVLYGLKKEGIRPIGVGGAVGNAHQIMYKSGNNRLLDPTTGVAAL-TFNKLM 202
QY 143 SLEISDEGNI-----TWTSFEVROFANVNHIGG 171
D 203 SGQVPPEQVFRVFKAKTIDSFDRVYRAVG--GG 237

Query Match 8.3%; Score 88.5; DB 16; Length 279;
Best Local Similarity 24.4%; Pred. No. 28;
Matches 53; Conservative 33; Mismatches 74; Indels 57; Gaps 12;

QY 3 KFLIIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-ATAAIEQSETID---PMK 50
D 30 RYLATAVVALAVAAALYARQDPEHFDEV---VLQCLDQAMSTVRRGSDGILPAC 86
QY 51 VPDHADFEHVGIVDFKGLAMENIAR-GLKQMKQGDANVKGEGIVKAHL-----104
D 87 QPELAELVGEDFGAVD---EPTLRVLATVAAANFAEYSSAVAYEAIEKSOHLNCGNL 143
QY 105 -----IGVHDDIVSMEYDLAYKLGD-----LHPTTHVTSIDQDFVAL 142
D 144 VFLMGVLYGLKKEGIRPIGVGGAVGNAHQIMYKSGNNRLLDPTTGVAAL-TFNKLM 202
QY 143 SLEISDEGNI-----TWTSFEVROFANVNHIGG 171
D 203 SGQVPPEQVFRVFKAKTIDSFDRVYRAVG--GG 237

RESULT 34
Q7W150 PRELIMINARY; PRT; 279 AA.
AC Q7W150;
DT 01-OCT-2003 (TRENBLrel. 25, Created)

```

```

DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative exported protein.
GN BP0840.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Doggett J.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640425; CAE40249.1; -.
KW Complete proteome.
SQ SEQUENCE 279 AA; 30292 MW; 0092B34E2CD15E26 CRC64;

Query Match 8.3%; Score 88.5; DB 16; Length 279;
Best Local Similarity 24.4%; Pred. No. 28;
Matches 53; Conservative 33; Mismatches 74; Indels 57; Gaps 12;

QY 3 KFLIIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-ATAAIEQSETID---PMK 50
D 30 RYLATAVVALAVAAALYARQDPEHFDEV---VLQCLDQAMSTVRRGSDGILPAC 86
QY 51 VPDHADFEHVGIVDFKGLAMENIAR-GLKQMKQGDANVKGEGIVKAHL-----104
D 87 QPELAELVGEDFGAVD---EPTLRVLATVAAANFAEYSSAVAYEAIEKSOHLNCGNL 143
QY 105 -----IGVHDDIVSMEYDLAYKLGD-----LHPTTHVTSIDQDFVAL 142
D 144 VFLMGVLYGLKKEGIRPIGVGGAVGNAHQIMYKSGNNRLLDPTTGVAAL-TFNKLM 202
QY 143 SLEISDEGNI-----TWTSFEVROFANVNHIGG 171
D 203 SGQVPPEQVFRVFKAKTIDSFDRVYRAVG--GG 237

RESULT 35
Q822J2 PRELIMINARY; PRT; 1393 AA.
AC Q822J2;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE DNA-directed RNA polymerase, beta' subunit.
GN RPOC OR CCA00690.
OS Chlamydomonada caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzaple E., Khouri H., Federova N.B., Carty H.A.,
RA Umavam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomonada caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the

```

```

RA Weber N., Grueninger D., Schmidheini T., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RL EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035528; CAB36839.1; -.
DR EMBL; AL161537; CAB78417.1; -.
DR FIR; T05244; T05244.
KW Hypothetical protein.
SQ
SQ SEQUENCE 2137 AA; 240107 MW; 0A3AFCB275C2CBC2 CRC64;

Query Match      8.3%; Score 88.5; DB 10; Length 2137;
Best Local Similarity 21.8%; Pred. No. 4.2e+02;
Matches 45; Conservative 34; Mismatches 82; Indels 45; Gaps 10

QY 35 DAIAAEQSEITIDPMKVPDHADK----FERHVGIVDFKGLAM--RNIEAR-GLKQMKRQ 87
DB 161 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 614 DSHFILELVQNADDNKYPEHVEPTLTILQKTGIIVLNNECGFMPENIRALCDVGSTKK 673
QY 88 GDANKVGEGEIVXAHLLIG-----VHDDIVSMYDIAY-KLGDLIPTT---HVISDIQDF 138
DB 161 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 674 GSGGVIGKGGFKGSFRVSDAPEIHSGHFHFKFISEQIGVILPTVVPFHDISSLGM 733
QY 139 VVALSLETSDEG---NITMTSFEVRQFANVNHIGLSILDPIF----- 179
DB 161 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 734 LSGRALHKDAGWNTCTILFRAIDSERTVNH-----EFMSDLHPSLLLFLRLQC 787
QY 180 -----GVLSDLVLTALPQDTVRKEMTKV 201
DB 161 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 788 IYVRNVLDSDLLVMRKEVSKNIVKV 813

RESULT 37
Q8FGF6 PRELIMINARY; PRT; 609 AA.
ID AC Q8FGF6;
AC Q8FGF6; PRELIMINARY; PRT; 609 AA.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter protein.
GN C2421.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=06:HI / CFT073 / ATCC 700928;
RX MEDLINE=23238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Licu S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
RL EMBL; AF016762; AAN80880.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0006166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_Arpase.
DR InterPro; IPR001140; ABC_TM_transpt.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER 1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER 2; 1.

```

```
KW Complete proteome.
SQ SEQUENCE 609 AA; 67758 MW; F7FB7AFF75E07378 CRC64;

Query Match      8.2%; Score 88; DB 16; Length 609;
Best Local Similarity 22.2%; Pred. No. 88;
Matches 43; Conservative 26; Mismatches 65; Indels 60; Gaps 6;

QY 6 LIAAVAFVAVSADP-----IHYKITEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH 61
DB LIAAVAFVAVSADP-----IHYKITEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH 61
DB LIAAVAFVAVSADP-----IHYKITEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH 61
DB LIAAVAFVAVSADP-----IHYKITEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH 61
QY 62 VGIYDFKXGELAMRNIEARGLKQMKRQGDANVKGBEGIVKAHLIGVHDDIVSMEYDLAYK 121
DB VGIYDFKXGELAMRNIEARGLKQMKRQGDANVKGBEGIVKAHLIGVHDDIVSMEYDLAYK 121
DB VGIYDFKXGELAMRNIEARGLKQMKRQGDANVKGBEGIVKAHLIGVHDDIVSMEYDLAYK 121
DB VGIYDFKXGELAMRNIEARGLKQMKRQGDANVKGBEGIVKAHLIGVHDDIVSMEYDLAYK 121
QY 122 LGDLHPTHTVSDIQDFVVALSLEISDEGNITMTSFVRQFANVNVN-----HIGLSILDPI 178
DB LGDLHPTHTVSDIQDFVVALSLEISDEGNITMTSFVRQFANVNVN-----HIGLSILDPI 178
DB LGDLHPTHTVSDIQDFVVALSLEISDEGNITMTSFVRQFANVNVN-----HIGLSILDPI 178
DB LGDLHPTHTVSDIQDFVVALSLEISDEGNITMTSFVRQFANVNVN-----HIGLSILDPI 178
QY 364 EGDGHALNHVSLTPPAASMSALVCAGSAGKTTVTKLMLR-YADPQQQSIGGVDIRRLT 422
DB EGDGHALNHVSLTPPAASMSALVCAGSAGKTTVTKLMLR-YADPQQQSIGGVDIRRLT 422
DB EGDGHALNHVSLTPPAASMSALVCAGSAGKTTVTKLMLR-YADPQQQSIGGVDIRRLT 422
DB EGDGHALNHVSLTPPAASMSALVCAGSAGKTTVTKLMLR-YADPQQQSIGGVDIRRLT 422
QY 179 FGVLSVLTAIFQD 192
DB FGVLSVLTAIFQD 192
DB FGVLSVLTAIFQD 192
DB FGVLSVLTAIFQD 192
QY 423 PEQLNSLISVVFDQ 436
DB PEQLNSLISVVFDQ 436
DB PEQLNSLISVVFDQ 436
DB PEQLNSLISVVFDQ 436

RESULT 38
Q897K6 PRELIMINARY; PRT; 485 AA.
AC Q897K6;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Uroporphyrinogen III synthase/methyltransferase (EC 4.2.1.75).
CN CTC00728
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA Becker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RA "The genome sequence of Clostridium tetani, the causative agent of
RA tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR EMBL; AE015938; AA055330.1;
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008168; F:methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0004852; F:uroporphyrinogen-III synthase activity; IEA.
DR GO; GO:0006783; P:heme biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000878; Cor/por_Mettransf.
DR InterPro; IPR003754; HEM4_synth.
DR InterPro; IPR003043; Uropor_Mettransf.
DR Pfam; PF02602; HEM4_1.
DR Pfam; PF00590; TP_methylase; 1.
DR PROSITE; PS00839; SUMT_1; 1.
DR PROSITE; PS00839; SUMT_1; 1.
KW Lysase; Methyltransferase; Transferrase; Complete proteome.
SQ SEQUENCE 485 AA; 54624 MW; B09D45F85570F747 CRC64;

Query Match      8.2%; Score 87.5; DB 16; Length 485;
Best Local Similarity 20.9%; Pred. No. 71;
Matches 44; Conservative 51; Mismatches 83; Indels 33; Gaps 10;

QY 27 BEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH-VGIYDFKXGELAMRNIEARGL---K 82
DB BEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH-VGIYDFKXGELAMRNIEARGL---K 82
DB BEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH-VGIYDFKXGELAMRNIEARGL---K 82
DB BEINKAIDDAIAIEQSETIDPMKVPDHPADKFERH-VGIYDFKXGELAMRNIEARGL---K 82
QY 262 KELREKSLDLGAQVTE---INAIKSRDELEKYEVLKDYK-FIVITSVNANILFDY 317
DB KELREKSLDLGAQVTE---INAIKSRDELEKYEVLKDYK-FIVITSVNANILFDY 317
DB KELREKSLDLGAQVTE---INAIKSRDELEKYEVLKDYK-FIVITSVNANILFDY 317
DB KELREKSLDLGAQVTE---INAIKSRDELEKYEVLKDYK-FIVITSVNANILFDY 317
QY 83 QMKRQGD-ANVKGBEGIV-----KAHLIGVHDDIVSMEYDLAYKGLDHPHTHTVSDIQ 136
DB QMKRQGD-ANVKGBEGIV-----KAHLIGVHDDIVSMEYDLAYKGLDHPHTHTVSDIQ 136
DB QMKRQGD-ANVKGBEGIV-----KAHLIGVHDDIVSMEYDLAYKGLDHPHTHTVSDIQ 136
DB QMKRQGD-ANVKGBEGIV-----KAHLIGVHDDIVSMEYDLAYKGLDHPHTHTVSDIQ 136
```

```
DB 318 LVKKEYDIRNTKGRFANVIGPATEKALKKRGKIKADIIAEFVAEDLFNELKP--VINKDDK 375
QY 137 DFV-----VALSLEISDEGNITMTSFVRQFANVNVNHHIGLSILDPIF-----GVL 182
DB 376 IIPVPHSKARQVLEALKDKG---CHVEEVYSYEVWKGVEVNLBIFDDVDIVLTSPSIV 432
QY 183 SDVLTAIFQDVTWKEMTKVLAPAFKRELEKN 213
DB 433 RNMDIMGLEKIKKTSISIGPITAKELDKN 463
DB 433 RNMDIMGLEKIKKTSISIGPITAKELDKN 463
DB 433 RNMDIMGLEKIKKTSISIGPITAKELDKN 463
DB 433 RNMDIMGLEKIKKTSISIGPITAKELDKN 463
RESULT 39
Q8IK83 PRELIMINARY; PRT; 1620 AA.
AC Q8IK83;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
CN PF14_0723.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368664;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Rairamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mangall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.J.;
RA "Genome sequence of the human malaria parasite Plasmodium
RA falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014828; AA037336.1;
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR002181; Fibrinogen C.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR004043; LCCL_dom.
DR InterPro; IPR008997; RicinB-like.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF03815; LCCL; 1.
DR SMART; SM00603; LCCL; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS00022; FA58C_3; 1.
DR PROSITE; PS00820; LCCL; 1.
DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 1620 AA; 184874 MW; 954B08097161F049 CRC64;

Query Match      8.2%; Score 87.5; DB 5; Length 1620;
Best Local Similarity 22.9%; Pred. No. 3.5e+02;
Matches 43; Conservative 37; Mismatches 53; Indels 55; Gaps 9;

QY 61 HVGIVDFKXGELAMRNIEARGLKQMKRQGDANV-----KGBEGIVKAHLIGVHDDIVSM 114
DB HVGIVDFKXGELAMRNIEARGLKQMKRQGDANV-----KGBEGIVKAHLIGVHDDIVSM 114
DB HVGIVDFKXGELAMRNIEARGLKQMKRQGDANV-----KGBEGIVKAHLIGVHDDIVSM 114
DB HVGIVDFKXGELAMRNIEARGLKQMKRQGDANV-----KGBEGIVKAHLIGVHDDIVSM 114
QY 810 HSDIIDNKGGLVNTIES-GMDHYVGSINNINIESISLNKNEKGLL-----DIIPE 858
DB HSDIIDNKGGLVNTIES-GMDHYVGSINNINIESISLNKNEKGLL-----DIIPE 858
DB HSDIIDNKGGLVNTIES-GMDHYVGSINNINIESISLNKNEKGLL-----DIIPE 858
DB HSDIIDNKGGLVNTIES-GMDHYVGSINNINIESISLNKNEKGLL-----DIIPE 858
QY 115 EYDLAYKGLD-----LHPTHTVSDIQDFVVAL-----SLSD 148
DB EYDLAYKGLD-----LHPTHTVSDIQDFVVAL-----SLSD 148
DB EYDLAYKGLD-----LHPTHTVSDIQDFVVAL-----SLSD 148
DB EYDLAYKGLD-----LHPTHTVSDIQDFVVAL-----SLSD 148
QY 859 EKEGNNIRRESSIFHHKTIYVSSLIEDCDLJLFLNQTSTFLEKGNIRNNKTKELKND 918
DB EKEGNNIRRESSIFHHKTIYVSSLIEDCDLJLFLNQTSTFLEKGNIRNNKTKELKND 918
DB EKEGNNIRRESSIFHHKTIYVSSLIEDCDLJLFLNQTSTFLEKGNIRNNKTKELKND 918
DB EKEGNNIRRESSIFHHKTIYVSSLIEDCDLJLFLNQTSTFLEKGNIRNNKTKELKND 918
QY 149 EGNITMTSFVRQFANVNVNHHIGLSILDPIFGVLSVLTAIFQDVT--KEMTKVLAPA- 205
DB EGNITMTSFVRQFANVNVNHHIGLSILDPIFGVLSVLTAIFQDVT--KEMTKVLAPA- 205
DB EGNITMTSFVRQFANVNVNHHIGLSILDPIFGVLSVLTAIFQDVT--KEMTKVLAPA- 205
DB EGNITMTSFVRQFANVNVNHHIGLSILDPIFGVLSVLTAIFQDVT--KEMTKVLAPA- 205
```

```
Db 919 DENWT-----VKQFHELISNL--MENIDAHGVDSVISIVQETIRIIEKTKKLPAD 971
QY 206 --FKRELE 211
Db 972 MLKKQIE 979

RESULT 40
Q8KV75 PRELIMINARY; PRT; 186 AA.
AC Q8KV75;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 60 kDa chaperonin (Fragment).
GN Cpn60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.I., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
RT throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437156; F:ATP binding; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; cpn60_TCF1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20155 MW; E78884004BACABED CRC64;

Query Match 8.1%; Score 87; DB 2; Length 186;
Best Local Similarity 24.6%; Pred. No. 22;
Matches 29; Conservative 23; Mismatches 32; Indels 34; Gaps 6;

QY 13 VASADPIHVDKITEINKAIDAI--EQTIDPMKVPDHPADKFER--HVGIV 65
Db 16 VTAGAPM-----DKRTDRAVAKVVENIAEAEV-----GDKFEKIEHVAKI 60

QY 66 DFKGELAMRNIEARGLKQKQGDANY---KGEGIVYKAHLIGVHDDIVSMYDILAY 120
Db 61 SANGDEAIGKLIAMQVRKKEGVIIVTEAKGTETV-----DVVEGMCQFDRGY 109

RESULT 41
Q9ZVB0 PRELIMINARY; PRT; 372 AA.
AC Q9ZVB0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 60 kDa chaperonin (Fragment).
GN F9K20.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,
RA Kremenetskaia I., Lueros J., Araujo R., Buehler E., Conway A.B.,
RA Dewar K., Feng J., Kim C., Li Y., Shinn P., Davis R.W., Ecker J.R.,
RA Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana full-length cDNA."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK118621; BAC43219.1; -
DR EMBL; BT008318; AAP37677.1; -
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR SMART; SM00028; TPR; 4.
KW Hypothetical protein.
SQ SEQUENCE 385 AA; 42579 MW; 6608867205554E48 CRC64;

Query Match 8.1%; Score 87; DB 10; Length 385;
Best Local Similarity 24.8%; Pred. No. 58;
Matches 40; Conservative 27; Mismatches 58; Indels 36; Gaps 8;
```

```
RT "Arabidopsis thaliana chromosome 1 BAC F9K20 sequence."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005679; AAC83019.1; -
DR FIR; G96818; G96818.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
SQ SEQUENCE 372 AA; 41053 MW; AFD052A81D1877A9 CRC64;

Query Match 8.1%; Score 87; DB 10; Length 372;
Best Local Similarity 24.8%; Pred. No. 55;
Matches 40; Conservative 27; Mismatches 58; Indels 36; Gaps 8;

QY 27 EINKAIDDAIAIEQSETIDPMKVPDHPADKFERHVG---IVDFKGLAMRNIEARGLKQ 83
Db 38 EEKDPALQQRKSSKQSVSV-PRKAPGLNTQFEKSGRSFIDIDFDERL--ENIRSALEQ 94

QY 84 MKRQ-----GDANVKGEGIVYKAHLIGVHDDIVSMYDILAYKLGDLHPT----- 128
Db 95 KTEVVEKFGFIDYDAPVKSQKTIGLTKRGV--GIAVVVGLVFAVGDFLPTGSDSPT 152

QY 129 --THVISDIQDFVVALSLEISDEGNITMTSFEVRQFANVN 167
Db 153 KNTTVVKN-----QISEEKATLQQ-RLKEFETTLN 182

RESULT 42
Q8GWV1 PRELIMINARY; PRT; 385 AA.
AC Q8GWV1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Atlg78915).
GN Atlg78915.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK118621; BAC43219.1; -
DR EMBL; BT008318; AAP37677.1; -
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR SMART; SM00028; TPR; 4.
KW Hypothetical protein.
SQ SEQUENCE 385 AA; 42579 MW; 6608867205554E48 CRC64;

Query Match 8.1%; Score 87; DB 10; Length 385;
Best Local Similarity 24.8%; Pred. No. 58;
Matches 40; Conservative 27; Mismatches 58; Indels 36; Gaps 8;
```

```
QY 27 EEINKAIDDAIAAEQSEITIDPMKVPDADKFEHVG---IVDFKGLAMRNTEARGLQ 83
Db 38 EEKDPALQQRKSSKQSVV-FRPAQGLNTQFEKSGRSFDIDFERL--ENIRSALEQ 94
QY 84 MKRQ-----GDANVKGEEGIVKAHLIGVHDDIVSMYDIAVYKLGDLHPT----- 128
Db 95 KKEVVEFGFDYDAPVKSDQKTIGLTKGV--GIAVVVGLVFPALGDFLFTGSDSPT 152
QY 129 --THVDSIQDFVVALSLEISDEGNITMTSFVVRQFANVN 167
Db 153 KNTTWN-----QISEEKATLQO-RLKEFFETILN 182

RESULT 43
ID O44999 PRELIMINARY; PRT; 399 AA.
AC O44999;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN W03D8.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Jones K., Graves T., Ozersky P.;
RT "The sequence of C. elegans cosmid W03D8."
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043702; AAK21491.1; -.
DR PIR; T32933; T32933.
DR WormPep; W03D8.5; C818325.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000535; MSP domain.
DR InterPro; IPR008962; PapD-like.
DR Pfam; PF00635; MSP domain; 1.
DR PROSITE; PS0202; MSP; 1.
KW Hypothetical protein.
SQ SEQUENCE 399 AA; 45411 MW; 050FE0780CB62190 CRC64;

Query Match 8.1%; Score 87; DB 5; Length 399;
Best Local Similarity 20.3%; Pred. No. 61; Indels 56; Gaps 7;
Matches 40; Conservative 38; Mismatches 63;

QY 11 AFVAYSADPIHYD-----KITEINKAIDDAIAAEQSEITIDPMKV 51
Db 44 SLVVVAAPVQDFRFAENQETTTQKSEEMGEVEEKKKKKKKEENKKAKEVETPEKE- 102
QY 52 PDHAKPFHHVGIYDFKGLAMRNTEARGLQMKRQGDANVKGEEGIVKAHLIGVHDDI 111
Db 103 ----EKIEKKKKKKKKKEEVEKPEKKEEKKKEEKKVGVNESRK-----IPDTN 153
QY 112 VSMYDIAVYKLGDLHPTTH-----VISDIQDFVVALSLEISDEGNITMTSFVVRQFANV 165
Db 154 TKGSHVSAVPPQQLFTPDCEQKLTITNTHDKKIMYKGLSD--NI---AFQ----- 201
```

```
QY 166 VNHIGLSILDPIFGVL 182
Db 202 -----LNPVFGVL 209
QY 202 -----LNPVFGVL 209

RESULT 44
ID Q8U2M6 PRELIMINARY; PRT; 562 AA.
AC Q8U2M6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein PF0807.
GN PF0807.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010197; AAL80931.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 562 AA; 64247 MW; 16EA3FE3946BB0FA CRC64;

Query Match 8.1%; Score 87; DB 17; Length 562;
Best Local Similarity 22.7%; Pred. No. 95;
Matches 51; Conservative 36; Mismatches 70; Indels 68; Gaps 12;

QY 22 YDKITEINKAIDDAIAAEQSEITIDPMKVPDADKFEHVG-IVD-----FKG----- 69
Db 63 YELTSLINFLIE-----EQNERLNFLPP--FLEFHRGIGKIEGQRAFFKGMVLSGV 114
QY 70 -----ELAMRNIARGLKQKQGDANV-KGEEGIVKAHLIGVHDDIVSMYDYL 118
Db 115 ANMKEGISESELALSNIESLEFKEDEGEKLNVDKLERVELVSAYEFLARKVNA 174
QY 119 AVKGLDHPHTHVISDIQDFVVALSLEISDEGNITMTSFVVRQFANVNHIGGLSI----- 174
Db 175 -----PLVLYSDENPFYI-----NVTFYGI-ARNFSEVILHIGNESFKVKV 216
QY 175 -----LDPIF--GV-----LSDLTAIFQDVTVRKEMTKVLAP 204
Db 217 VEGEFSLDYIFVSPGVYIAYAEAGNVKSNVIRINVTKIPTRIAP 261

RESULT 45
ID Q9AZS0 PRELIMINARY; PRT; 1640 AA.
AC Q9AZS0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Tail protein.
GN ORF52.
OS Bacteriophage b11286.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=151536;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21113149; PubMed=11160885;
RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RT "Analysis of six prophages in Lactococcus lactis IIL403: different
RT genetic structure of temperate and virulent phage populations."
RL Nucleic Acids Res. 29:644-651(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323669; AAK08339.1; -.

```

```
SQ SEQUENCE 1640 AA; 176321 MW; CEE59A7A47B8F18 CRC64;
Query Match 8.1%; Score 87; DB 9; Length 1640;
Best Local Similarity 20.7%; Pred. No. 3.9e+02;
Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;

QY 16 SADPIHYDKITEINKAIDDAIAAQSEETIDPMKVPDHDADKFERHVGIVDFKGLAMRN 75
DB 1142 SMSKAQYDEIVKNAQKQDRTISAACKQQT-----EVTDKAQK--THDKTVELANSKADKN 1195
QY 76 IEA-----RGLQMKRQGDANYKGBEGIVK-----AHLLI----- 105
DB 1196 VKAAAKEQGETVEQYTKGFKDSNLNSFIDGINGVNLFLHKGWNGHVSCLKGFATGTR 1255
QY 106 GVHDDIVSMEYDLAYKGLDHPHTHVIS-----DIQDFVALSL---EISDEGNITWTS 156
DB 1256 GLAQDETALVGBEGFELAH-HPSRGIFAVGQQQPEIRNLKAGTSILPHSMSKE-FLSLT- 1312
QY 157 FEVRQFANVNVHIGLSILDPIFGVLSDLVLTAFQDVTVRKMTKVLA 203
DB 1313 -----ANLPAHADGVS-----GFLSDALGWV--KSTYKDVTSVIS 1345

RESULT 46
Q9CF57 PRELIMINARY; PRT; 1640 AA.
AC Q9CF57
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Prophage p13 protein 14.
GN P1314 OR L13138.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1366;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=111403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006370; AAK05486.1; -.
DR PIR; D86798; D86798.
KW Complete proteome.
SQ SEQUENCE 1640 AA; 176320 MW; CEE59A7A47B8F18 CRC64;

Query Match 8.1%; Score 87; DB 16; Length 1640;
Best Local Similarity 20.7%; Pred. No. 3.9e+02;
Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;

QY 16 SADPIHYDKITEINKAIDDAIAAQSEETIDPMKVPDHDADKFERHVGIVDFKGLAMRN 75
DB 1142 SMSKAQYDEIVKNAQKQDRTISAACKQQT-----EVTDKAQK--THDKTVELANSKADKN 1195
QY 76 IEA-----RGLQMKRQGDANYKGBEGIVK-----AHLLI----- 105
DB 1196 VKAAAKEQGETVEQYTKGFKDSNLNSFIDGINGVNLFLHKGWNGHVSCLKGFATGTR 1255
QY 106 GVHDDIVSMEYDLAYKGLDHPHTHVIS-----DIQDFVALSL---EISDEGNITWTS 156
DB 1256 GLAQDETALVGBEGFELAH-HPSRGIFAVGQQQPEIRNLKAGTSILPHSMSKE-FLSLT- 1312
QY 157 FEVRQFANVNVHIGLSILDPIFGVLSDLVLTAFQDVTVRKMTKVLA 203
DB 1313 -----ANLPAHADGVS-----GFLSDALGWV--KSTYKDVTSVIS 1345

RESULT 47
Q7TZ25 PRELIMINARY; PRT; 280 AA.
ID Q7TZ25
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Conserved hypothetical protein.
CN MB1518C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RX Garnier T., Bigmelter K., Camus J.-C., Medina N., Mansoor H.,
RA Fryer M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248339; CAD96185.1; -.
KW Complete proteome.
SQ SEQUENCE 280 AA; 31664 MW; 2110B7D7A51B4EE7 CRC64;

Query Match 8.1%; Score 86; DB 16; Length 280;
Best Local Similarity 28.3%; Pred. No. 46;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;

QY 33 IDDAIAIEQSETIDPMKVPDHDADKFERHVGIVDFKGLAMRNIEARGLQMKR----- 86
DB 128 LDPVAADVADLIQATDLKVDVEPLIERYGRGRGKAAARALDLVDGGAQSPKETWLRLL 187
QY 87 -----QGDANVKGEEGIVKAHLIGVHDDIVSMEYDLAYKGLDHPHTT--HVISDI 135
DB 188 LIRAGFRPQTQIAVRNENWGAHLQDQWQDIKVAEYD-----GDHLITSRYHYEKDI 242

RESULT 48
Q91432 PRELIMINARY; PRT; 303 AA.
ID Q91432
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Probable transcriptional regulator.
GN PA1312.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.E., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 405:959-964(2000).
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL; AE004560; AAG04701.1; -.
DR PIR; D83482; D83482.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000847; HTH_LysR.
DR InterPro; IPR005119; LysR_subst.
DR Pfam; PF0126; HTH_1; 1.
DR Pfam; PF03466; LysR_substrate; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
```



```
KW DNA-binding; Transcription regulation; Complete proteome.
.SQ SEQUENCE 303 AA; 33062 MW; F8C201582A0BF94 CRC64;

Query Match      8.1%; Score 86; DB 16; Length 303;
Best Local Similarity 28.6%; Pred. No. 51;
Matches 28; Conservative 19; Mismatches 39; Indels 38; Gaps 4;

QY 24 KITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDFKGLAMRNIEARGLKQ 83
Db 71 KVLDEARLAIDARA-----EKG-----ELRGLRVTITQYGLRQ 106
QY 84 MKRQGDANVKEEGIVKAHLIGV-----HDDIVSMEYDLAVKGLDHPHTTHVISDIQ 136
Db 107 L-----VPALQAFARLHPALQVLTSSLIHADLIGERFDVAIRGLRSDSTHHAVQLA 159
QY 137 DFVV 140
Db 160 SFEV 163

RESULT 49
P71763 PRELIMINARY; PRT; 339 AA.
AC P71763;
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein Rv1482c.
GN Rv1482c OR MT1529 OR M7C1277.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO THE GRPE FAMILY.
DR EMBL; Z79701; CAB02032.1; -.
DR EMBL; AE007022; AAK45794.1; ALT_INIT.
DR PIR; E70710; E70710.
DR TIGR; MT1529; -.
DR Tuberculin; Rv1482c; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 339 AA; 38290 MW; A0526C082B219D18 CRC64;

Query Match      8.1%; Score 86; DB 16; Length 339;
Best Local Similarity 28.3%; Pred. No. 59;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;

QY 33 IDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDFKGLAMRNIEARGLKQMKR----- 86

KW DNA-binding; Transcription regulation; Complete proteome.
.SQ SEQUENCE 303 AA; 33062 MW; F8C201582A0BF94 CRC64;

Query Match      8.1%; Score 86; DB 16; Length 303;
Best Local Similarity 28.6%; Pred. No. 51;
Matches 28; Conservative 19; Mismatches 39; Indels 38; Gaps 4;

QY 24 KITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDFKGLAMRNIEARGLKQ 83
Db 71 KVLDEARLAIDARA-----EKG-----ELRGLRVTITQYGLRQ 106
QY 84 MKRQGDANVKEEGIVKAHLIGV-----HDDIVSMEYDLAVKGLDHPHTTHVISDIQ 136
Db 107 L-----VPALQAFARLHPALQVLTSSLIHADLIGERFDVAIRGLRSDSTHHAVQLA 159
QY 137 DFVV 140
Db 160 SFEV 163

RESULT 50
Q06166 PRELIMINARY; PRT; 1661 AA.
AC Q06166;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Mature PARASITE-infected erythrocyte surface antigen (Antigenic
DE protein PFEMP2).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RX MEDLINE=92158014; PubMed=1741020;
RA Coppel R.L.;
RT "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT binds human erythrocyte protein 4.1";
RL Mol. Biochem. Parasitol. 50:335-347 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RA Kun J.F.J., Waller K.L., Coppel R.L.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 797-850 FROM N.A.
RX MEDLINE=93122844; PubMed=1478701;
RA Saul A., Venganeh F., Howard R.J.;
RT "Conservation of repeating structures in the pFEMP2/MESA protein of
RT Plasmodium falciparum";
RL Immunol. Cell Biol. 70:353-355 (1992).
DR EMBL; AF056936; AAC13303.1; -.
DR EMBL; S52458; AAE24869.1; -.
DR PIR; B56657; B56657.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DnaJ_2; 1.
KW Antigen.
SQ SEQUENCE 1661 AA; 195479 MW; AF340527D85A9D29 CRC64;

Query Match      8.1%; Score 86; DB 5; Length 1661;
Best Local Similarity 20.7%; Pred. No. 4.8e+02;
Matches 46; Conservative 41; Mismatches 67; Indels 68; Gaps 10;

QY 15 VSADIHVDKITEINKAIDDAIAAEQSETIDPMKVP-----DHADKFERHVGIVDFK 69
Db 604 VNGPILIEEVKEIKKQVGEIKKE-NDTEGNDKVGKEIIESEVKEIKKHV----- 655
QY 70 ELAMENIEARGLKQMKRQGDANVKEEGI-----VKAHLIGVHD-----D 110
Db 656 -----EEGKENNTEGNDKVGPEIIEEVKEIKKHVEEGIKENDTESKDKVIGQE 707
QY 111 IVSMEYDLAVKGLDHPHTTHVISDIQDFVVALSLEISDEGNITMTSPFVRQFANVNHIG 170
Db 708 IITEVKEGIKEND-----TENKDKVIGPEM-IITE-----VKKEIEKQEKGNKEN 753
QY 171 GLSILDPFGVLSVLTAFQDTVRKENTKVLAPAFKRELEK 212
Db 754 ILEIKDIVIG-----QEVIIIEVKVI-----KKVKEK 781

RESULT 51
Q8KV93
```

```
ID AC Q8KV93 PRELIMINARY; PRT; 186 AA.
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 60 kDa chaperonin (Fragment).
GN CPN60
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF437131; AAM49392.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/ICP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR ATP-binding; Chaperone.
FT NON TER 1 186
FT SEQUENCE 186 AA; 19733 MW; 8BAS55908561D40A CRC64;
SQ
Query Match
Best Local Similarity 8.0%; Score 85.5; DB 2; Length 186;
Matches 41; Conservative 26; Mismatches 64; Indels 39; Gaps 8;
QY 13 VAVSADPIHYDKITEINKAIDAAIAEQSETIDPMKVPDHDADKFERHVGIVDPKGEIA 72
DB 16 VTAGASPI-----DLKRGIDKAVAAVVAELKAHSEVGGDDYSKVEQ-VGTVSANDSA 67
QY 73 MRNIEARGLKQMKRQGDANY---KGEIGIVKAHLIGVHDDIVSMYEDLAYKLGDLHPTT 129
DB 68 IGKLIADAMSKVKKGDTITVEEAKGTETEVK-----VEGQFDRGY---ISPYF 114
QY 130 HVLSIDQDFVALSLEISDEGNITWTSFEVRQFANVNVHIGLSILDP 178
DB 115 MTNSD-----KMEASLSPVYLITDKKISSMKDL-----LPILPEI 150
RESULT 52
Q8KV93 PRELIMINARY; PRT; 186 AA.
ID Q8KV93
AC Q8KV93
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 60 kDa chaperonin (Fragment).
GN CPN60
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF437131; AAM49392.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/ICP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCF1; 1.
FT NON TER 1 186
FT SEQUENCE 186 AA; 19733 MW; 8BAS55908561D40A CRC64;
SQ
Query Match
Best Local Similarity 8.0%; Score 85.5; DB 2; Length 186;
Matches 41; Conservative 26; Mismatches 64; Indels 39; Gaps 8;
QY 13 VAVSADPIHYDKITEINKAIDAAIAEQSETIDPMKVPDHDADKFERHVGIVDPKGEIA 72
DB 16 VTAGASPI-----DLKRGIDKAVAAVVAELKAHSEVGGDDYSKVEQ-VGTVSANDSA 67
QY 73 MRNIEARGLKQMKRQGDANY---KGEIGIVKAHLIGVHDDIVSMYEDLAYKLGDLHPTT 129
DB 68 IGKLIADAMSKVKKGDTITVEEAKGTETEVK-----VEGQFDRGY---ISPYF 114
QY 130 HVLSIDQDFVALSLEISDEGNITWTSFEVRQFANVNVHIGLSILDP 178
DB 115 MTNSD-----KMEASLSPVYLITDKKISSMKDL-----LPILPEI 150
```

```
FT NON TER 1 1
FT NON TER 186
SQ SEQUENCE 186 AA; 19772 MW; 48F171D1BB55CB17 CRC64;
Query Match
Best Local Similarity 8.0%; Score 85.5; DB 2; Length 186;
Matches 38; Conservative 30; Mismatches 64; Indels 37; Gaps 7;
QY 13 VAVSADPIHYDKITEINKAIDAAIAEQSETIDPMKVPDHDADKFERHVGIVDPKGEIA 72
DB 16 VTAGANPW-----DLKRGIDKAVAAVVAELKAHSEVGGDDYSKVEQ-VGTVSANDAT 67
QY 73 MRNIEARGLKQMKRQGDANY---KGEIGIVKAHLIGVHDDIVSMYEDLAYKLGDLHPTT 129
DB 68 IGKLIADAMSKVKKGDTITVEEAKGTETEVK-----VEGQFDRGY---ISPYF 114
QY 130 HVLSIDQDFVALSLEISDEGNITWTSFEVRQFANVNVHIGLSILDP 178
DB 115 MTDSKMEAVL-----DNASVLITDKKISSMKDL-----MFILEPI 150
RESULT 53
Q86YC9 PRELIMINARY; PRT; 263 AA.
ID Q86YC9
AC Q86YC9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042924; AAH42924.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 263 AA; 29327 MW; E9A8FAA1C6A9CB00 CRC64;
Query Match
Best Local Similarity 8.0%; Score 85.5; DB 4; Length 263;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;
QY 20 IHYDKI--TEINKAIDAAIAEQSETIDPMK-----VPDADKFERHV-----GIVDF 67
DB 31 LHRDTTQTTLQPSVKDGLIVYEDSLVAKVKGHLVWDEADKAPTNTVCILKTLVE- 89
QY 68 KGEIAM---RNIEARGLKQMKRQGDANYKGEIGIVKAH-----LLIGVHDDIVSMYEDLA 119
DB 90 NGEMLADGRIVA-----NSANVNGENVVVIHPDFRMIVLANRPGFPLGNDFF 140
QY 120 YKLGDLHPTTHISD-----IQDFVALS-----IQDFVALS--LEISDEGNITWTS 156
DB 141 GTLGDI--SCHAVDNPXPHSEMLEMLRQYGNVPPEPIQLKLVAAFGELSLADQGIINY-P 198
QY 157 FEVRQFANVNVHIGLSILDPFGVLSVLTALFQ-DTVRKEMTKVL 202
DB 199 YSTREVNIVKHLQKF-----PTEG-LSSVVRNVDFDSYNNDMREIL 240
RESULT 54
Q7Y444 PRELIMINARY; PRT; 264 AA.
ID Q7Y444
AC Q7Y444;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gp22 prohead core scaffold protein.
OS Enterobacteria phage RB49.
```


01-OCT-2002 (TrEMBLrel. 22, Created)
01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
Hypothetical protein (Fragment).
OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_Taxid=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RA Bloesker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the ENBL/GenBank/DBJ databases.
DR EMBL; AL834181; CAD38878.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR SMART; SM00382; AAA; 2.
KW Hypothetical protein; ATP-binding.
FT NON_TER 1
SQ SEQUENCE 745 AA; 83522 MW; F8B818ED30CCC38 CRC64;

Query Match 8.0%; Score 85.5; DB 4; Length 745; -
Best Local Similarity 22.0%; Pred. No. 1.8e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHYDKI--TEINKAIDDAIAIEQSITIDPMK-----VPDHADKPERHV-----GIVDF 67
DB 513 LHRDITVQTLIQPSVKDGLIYEDSPLVKAVKLGHLVWDEADKAPHTVTCILKTLVE- 571
QY 68 KGEIAM--RNTEARGLKQMRQGDANVKGEEGIVKAH-----LLIGVHDDIVSMYEDLA 119
DB 572 NGEMLADGRIVA-----NSANVNGRENVVVHPDFRMIVLANRPGFFLNGDFF 622
QY 120 YKLGDLHPHTHVISD-----IQDFVVALS--LISIEGNIWTIS 156
DB 623 GTLGIF--SCHAVDNPKEPHELEMLRQGVNPEPIQLKLVAAFGSLRSLADQGIINY-P 680
QY 157 FEVROFANVNHIGSLSDIFGVLSDLVLTAFQ-DTVKEMTKVL 202
DB 681 YSTREWNIVKHLQKF----PTEG-LSGVNVPDFDSYNNDRIL 722

RESULT 57
Q7UYX3 PRELIMINARY; PRT; 807 AA.
ID Q7UYX3
AC Q7UYX3
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Copper-transporting ATPase (EC 3.6.3.4).
GN RB334.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_Taxid=117;
[1]
SEQUENCE FROM N.A.
RP STRAIN=1;
RC MEDLINE=22735913; PubMed=12835416;
RX Gloeckner P.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
RL EMBL; BX294133; CAD71518.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 807 AA; 85737 MW; 40D9588BE29484C4 CRC64;

Query Match 8.0%; Score 85.5; DB 16; Length 807;
Best Local Similarity 24.3%; Pred. No. 2e+02;

```
RESULT 59
Q72619 PRELIMINARY; PRT; 1062 AA.
AC Q72619;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eve;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Cay L.J., Hulyk S.W.,
RA Villalón D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eve;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053674; AAH53674.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 1062 AA; 119150 MW; E629CB68B018E21 CRC64;

Query Match 8.0%; Score 85.5; DB 4; Length 1062;
Best Local Similarity 22.0%; Pred. No. 2.9e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

Qy 20 IHVDKI--TEINKAIDDAIAIEQSEITIDPMK-----VPDHADKFERHV-----GIVDF 67
Db 830 LHRDTTQTTLQPSVKGDLIVVEDSPLVKAVKGLHVLVVDKAPNTVTCILKTLVE- 888
Qy 68 KGEIAM--RNIEARGLKQMKRGDANVKGEGIVKVAH-----LLGVHDDIVSMYDIA 119
Db 889 NGEMILADGRIVA-----NSANVNGRNVVWHPDFRMVILANRPGPFLLGNDFP 939
Qy 120 YKLGDLHPHTHVISD-----IQDFWALS--LEISDEGNTMTS 156
Db 940 GTLGDIIF-SCHAVNDPKPHESELMKRGVGNVPEPIQLKLVAFGELSLADQGIINY-P 997
Qy 157 FEVRQFQFNVNHHGGLSILDPFGVLSVDLTAIFQ-DTVKEMTKVL 202
Db 998 YSTREVVNIVKHLQKF----PTEG-LSSVVNRVDFDYSNNDMREIL 1039

RESULT 60
Q91250 PRELIMINARY; PRT; 1154 AA.
AC Q91250
AC Q91250;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998)
DR EMBL; AB011136; BAA25490.1; -.
DR FIR; T00335; T00335.
DR InterPro; IPR002035; VWF_A.
DR SMART; SMO0327; VWA; 1.
```

```
DR PROSITE; PS50234; VWFA; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1441 AA; 162836 MW; 360A40ECA4E3F054D CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 4.4e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHVDKI--TEEINKAIDDAIAEQSETIDPMK-----VPDHADKFERHV-----GIVDF 67
DB 343 LHRDITVQTLTQPSVKQGLVVEDSFLVKAVALGHILVVDEADKAPTNTVILKTLVE- 401
QY 68 KGLAM---RNIEARGLKMKRGDANVKGEGIVKAH-----LIGVHDDIVSMEYDLA 119
DB 402 NGEMILADGERIVA-----NSANVNGRENVVVIHPDFRMVLANRPGFPFLGNDF 452
QY 120 YKGLDHLPTHVLSID-----IDPFVVALS---LEISDEGNITWTS 156
DB 453 GTLGDIIF-SCHAYDNPKPHEMLRQYGNVPEPIQLKLVAAFGSLRLADQGIINY-P 510
QY 157 FEVRQFANVNVHIGGLSILDPFVGLSDVLTAFQ-DTVRKEMTKVL 202
DB 511 YSTREVVNVKHLQKF-----PTSG-LSSVVRNVFDFDSYNNDRREIL 552

RESULT 62
QY1ZS1 ID Q91ZS1 PRELIMINARY; PRT; 2238 AA.
AC Q91ZS1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MASS1.2.
GN MUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21432018; PubMed=11545713;
RA Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;
RT "A novel gene causing a mendelian audiogenic mouse epilepsy.";
RL Neuron 31:537-544(2001).
DR EMBL; AF405693; AAL06013.1; -.
DR MGD; MGI:1274784; Mass1.
GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 12.
DR SMART; SM00237; Calx-beta; 11.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00879; ODR DC 2.2; 1.
SQ SEQUENCE 2238 AA; 244054 MW; AFE2B4F8A1BDAF58 CRC64;

Query Match
Best Local Similarity 21.4%; Pred. No. 7.9e+02;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAEQSETIDPMKVPDHADKFERHV-----GIVDFKGLAMRNIEA 78
DB 1966 TVTVNLANDNVAGIVSFQTASRSVIGHEGMLQFHVVRTPPGRGNTVNVKVGQNLV 2025
QY 79 RGLKMKRQGDANVKG-----EGIVKAHLIGVHDDIVSME-----YDLAYKGLDLH 126
DB 2026 -----NFANFTGQLFFSEGTINKTIFVHLLDDNIPKEKVVQVLYDV--KTQGS 2074
QY 127 PTHVISIDQFVVALSLEISDEG---NITWTS-FEVRQFANVV-----NHIGGLSIL 175

DR PROSITE; PS50234; VWFA; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1441 AA; 162836 MW; 360A40ECA4E3F054D CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 4.4e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHVDKI--TEEINKAIDDAIAEQSETIDPMK-----VPDHADKFERHV-----GIVDF 67
DB 343 LHRDITVQTLTQPSVKQGLVVEDSFLVKAVALGHILVVDEADKAPTNTVILKTLVE- 401
QY 68 KGLAM---RNIEARGLKMKRGDANVKGEGIVKAH-----LIGVHDDIVSMEYDLA 119
DB 402 NGEMILADGERIVA-----NSANVNGRENVVVIHPDFRMVLANRPGFPFLGNDF 452
QY 120 YKGLDHLPTHVLSID-----IDPFVVALS---LEISDEGNITWTS 156
DB 453 GTLGDIIF-SCHAYDNPKPHEMLRQYGNVPEPIQLKLVAAFGSLRLADQGIINY-P 510
QY 157 FEVRQFANVNVHIGGLSILDPFVGLSDVLTAFQ-DTVRKEMTKVL 202
DB 511 YSTREVVNVKHLQKF-----PTSG-LSSVVRNVFDFDSYNNDRREIL 552

RESULT 62
QY1ZS1 ID Q91ZS1 PRELIMINARY; PRT; 2238 AA.
AC Q91ZS1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MASS1.2.
GN MUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21432018; PubMed=11545713;
RA Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;
RT "A novel gene causing a mendelian audiogenic mouse epilepsy.";
RL Neuron 31:537-544(2001).
DR EMBL; AF405693; AAL06013.1; -.
DR MGD; MGI:1274784; Mass1.
GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 9.
DR SMART; SM00237; Calx-beta; 10.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00879; ODR DC 2.2; 1.
SQ SEQUENCE 2238 AA; 244054 MW; AFE2B4F8A1BDAF58 CRC64;

Query Match
Best Local Similarity 21.4%; Pred. No. 7.9e+02;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAEQSETIDPMKVPDHADKFERHV-----GIVDFKGLAMRNIEA 78
DB 1966 TVTVNLANDNVAGIVSFQTASRSVIGHEGMLQFHVVRTPPGRGNTVNVKVGQNLV 2025
QY 79 RGLKMKRQGDANVKG-----EGIVKAHLIGVHDDIVSME-----YDLAYKGLDLH 126
DB 2026 -----NFANFTGQLFFSEGTINKTIFVHLLDDNIPKEKVVQVLYDV--KTQGS 2074
QY 127 PTHVISIDQFVVALSLEISDEG---NITWTS-FEVRQFANVV-----NHIGGLSIL 175
```

```
DR PROSITE; PS50234; VWFA; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1441 AA; 162836 MW; 360A40ECA4E3F054D CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 4.4e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHVDKI--TEEINKAIDDAIAEQSETIDPMK-----VPDHADKFERHV-----GIVDF 67
DB 343 LHRDITVQTLTQPSVKQGLVVEDSFLVKAVALGHILVVDEADKAPTNTVILKTLVE- 401
QY 68 KGLAM---RNIEARGLKMKRGDANVKGEGIVKAH-----LIGVHDDIVSMEYDLA 119
DB 402 NGEMILADGERIVA-----NSANVNGRENVVVIHPDFRMVLANRPGFPFLGNDF 452
QY 120 YKGLDHLPTHVLSID-----IDPFVVALS---LEISDEGNITWTS 156
DB 453 GTLGDIIF-SCHAYDNPKPHEMLRQYGNVPEPIQLKLVAAFGSLRLADQGIINY-P 510
QY 157 FEVRQFANVNVHIGGLSILDPFVGLSDVLTAFQ-DTVRKEMTKVL 202
DB 511 YSTREVVNVKHLQKF-----PTSG-LSSVVRNVFDFDSYNNDRREIL 552

RESULT 62
QY1ZS1 ID Q91ZS1 PRELIMINARY; PRT; 2238 AA.
AC Q91ZS1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MASS1.2.
GN MUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21432018; PubMed=11545713;
RA Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;
RT "A novel gene causing a mendelian audiogenic mouse epilepsy.";
RL Neuron 31:537-544(2001).
DR EMBL; AF405693; AAL06013.1; -.
DR MGD; MGI:1274784; Mass1.
GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 12.
DR SMART; SM00237; Calx-beta; 11.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00879; ODR DC 2.2; 1.
SQ SEQUENCE 2238 AA; 244054 MW; AFE2B4F8A1BDAF58 CRC64;

Query Match
Best Local Similarity 21.4%; Pred. No. 7.9e+02;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAEQSETIDPMKVPDHADKFERHV-----GIVDFKGLAMRNIEA 78
DB 1966 TVTVNLANDNVAGIVSFQTASRSVIGHEGMLQFHVVRTPPGRGNTVNVKVGQNLV 2025
QY 79 RGLKMKRQGDANVKG-----EGIVKAHLIGVHDDIVSME-----YDLAYKGLDLH 126
DB 2026 -----NFANFTGQLFFSEGTINKTIFVHLLDDNIPKEKVVQVLYDV--KTQGS 2074
QY 127 PTHVISIDQFVVALSLEISDEG---NITWTS-FEVRQFANVV-----NHIGGLSIL 175

DR PROSITE; PS50234; VWFA; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1441 AA; 162836 MW; 360A40ECA4E3F054D CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 4.4e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHVDKI--TEEINKAIDDAIAEQSETIDPMK-----VPDHADKFERHV-----GIVDF 67
DB 343 LHRDITVQTLTQPSVKQGLVVEDSFLVKAVALGHILVVDEADKAPTNTVILKTLVE- 401
QY 68 KGLAM---RNIEARGLKMKRGDANVKGEGIVKAH-----LIGVHDDIVSMEYDLA 119
DB 402 NGEMILADGERIVA-----NSANVNGRENVVVIHPDFRMVLANRPGFPFLGNDF 452
QY 120 YKGLDHLPTHVLSID-----IDPFVVALS---LEISDEGNITWTS 156
DB 453 GTLGDIIF-SCHAYDNPKPHEMLRQYGNVPEPIQLKLVAAFGSLRLADQGIINY-P 510
QY 157 FEVRQFANVNVHIGGLSILDPFVGLSDVLTAFQ-DTVRKEMTKVL 202
DB 511 YSTREVVNVKHLQKF-----PTSG-LSSVVRNVFDFDSYNNDRREIL 552

RESULT 62
QY1ZS1 ID Q91ZS1 PRELIMINARY; PRT; 2238 AA.
AC Q91ZS1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MASS1.2.
GN MUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21432018; PubMed=11545713;
RA Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;
RT "A novel gene causing a mendelian audiogenic mouse epilepsy.";
RL Neuron 31:537-544(2001).
DR EMBL; AF405693; AAL06013.1; -.
DR MGD; MGI:1274784; Mass1.
GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 12.
DR SMART; SM00237; Calx-beta; 11.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00879; ODR DC 2.2; 1.
SQ SEQUENCE 2238 AA; 244054 MW; AFE2B4F8A1BDAF58 CRC64;

Query Match
Best Local Similarity 21.4%; Pred. No. 7.9e+02;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAEQSETIDPMKVPDHADKFERHV-----GIVDFKGLAMRNIEA 78
DB 1966 TVTVNLANDNVAGIVSFQTASRSVIGHEGMLQFHVVRTPPGRGNTVNVKVGQNLV 2025
QY 79 RGLKMKRQGDANVKG-----EGIVKAHLIGVHDDIVSME-----YDLAYKGLDLH 126
DB 2026 -----NFANFTGQLFFSEGTINKTIFVHLLDDNIPKEKVVQVLYDV--KTQGS 2074
QY 127 PTHVISIDQFVVALSLEISDEG---NITWTS-FEVRQFANVV-----NHIGGLSIL 175
```

[illegible]

AC Q8KVQ2; 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.,
RT Extensive profiling of a complex microbial community by high-throughput sequencing.
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF437042; AAM49303.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCPI-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCPI; 1.
KW ATP-binding; Chaperone.
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 20151 MW; 7BE9B5B3329114D CRC64;

Query Match 8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;
QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSEITIDPMKVPDHDADKFERHVGIVDFKGLA 72
DB 16 VAAGNPM-----DLKRGIDKAVAKVVEIGIKAAQAEVGDDEFDKIE-NVARISANDSV 67
QY 73 MRNIEARGLKQKRGDANVKGEIGIVKAHLILGVHDDIV-SMEYDLAY 120
DB 68 IGLIAEAMKVKRDGVITVEDAKG-----TDTYVDVVEGMQFDRGY 109

Query Match 8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;
QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSEITIDPMKVPDHDADKFERHVGIVDFKGLA 72
DB 16 VAAGNPM-----DLKRGIDKAVAKVVEIGIKAAQAEVGDDEFDKIE-NVARISANDSV 67
QY 73 MRNIEARGLKQKRGDANVKGEIGIVKAHLILGVHDDIV-SMEYDLAY 120
DB 68 IGLIAEAMKVKRDGVITVEDAKG-----TDTYVDVVEGMQFDRGY 109

RESULT 68

Q8KHQ8 PRELIMINARY; PRT; 186 AA.
AC Q8KHQ8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.,
RT Extensive profiling of a complex microbial community by high-throughput sequencing.
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF436897; AAM49158.1; -
DR EMBL; AF437007; AAM49268.1; -
DR EMBL; AF437129; AAM49390.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCPI-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCPI; 1.
KW ATP-binding; Chaperone.
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 20201 MW; 7BE9B5B26290CEF CRC64;

Query Match 8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;
QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSEITIDPMKVPDHDADKFERHVGIVDFKGLA 72
DB 16 VAAGNPM-----DLKRGIDKAVAKVVEIGIKAAQAEVGDDEFDKIE-NVARISANDSV 67
QY 73 MRNIEARGLKQKRGDANVKGEIGIVKAHLILGVHDDIV-SMEYDLAY 120
DB 68 IGLIAEAMKVKRDGVITVEDAKG-----TDTYVDVVEGMQFDRGY 109

RESULT 69

Q9UZZ6 PRELIMINARY; PRT; 292 AA.
AC Q9UZZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PYRAB10690.
GN PYRAB10690 OR PAB0711.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248286; CAB49980.1; -
DR PIR; G75084; G75084.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 292 AA; 31578 MW; AE410B3578BFA8C CRC64;

Query Match 8.0%; Score 85; DB 17; Length 292;
Best Local Similarity 20.7%; Pred. No. 59;
Matches 44; Conservative 31; Mismatches 62; Indels 76; Gaps 7;

QY 29 INKAIDD-----AIAAEQSEITIDPMKVPDHDADKFERHVGIVDFK---- 68
DB 11 INKVIENPKVIIFPLIVLIPSVPLAFLEKESNIKPLNFEEAGVILIEKHGAISDMKLPNL 70
QY 69 -----GELAMRNTA-----RGLQMKRGDANVKGEIGIVKAHLILGVHDDIVSMEX 116
DB 71 KLLLVGLQLLSAVQYSIIHYVKTGSTNGEAFKGLGVNIFLL----- 118
QY 117 DLAYKLGDLHPTTHVISDIDQFVVALSLEISDEGNITMTSFEVRQF-ANVVNHIGGL--S 173
DB 119 -----NVISTL-----IVLVAFLIAIFPPMIIVIGGILES 149
QY 174 ILDFIGVLSDLTAIFQDITVRKEMTKVLAPAF 206
DB 150 TGAVIFGILLVILSLIGSFAIGMTSIVPEAY 182

RESULT 70

Q8RDZ2 PRELIMINARY; PRT; 401 AA.
AC Q8RDZ2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ABC transporter permease protein.
GN FNI349.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;


```
DE Homoioisocitrate dehydrogenase.
GN HiCD.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococchi; Therrmales; Thermaceae;
OC Thermus.
ON NCBI_TaxID=274;
RX [1]_TaxID=274;
RN SEQUENCE FROM N.A.
RP Miyazaki J., Kobashi N., Nishiyama M., Yamane H.;
RT "Characterization of homoioscitrate dehydrogenase from Thermus
RL thermophilus HB27.";
RM Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RW EMBL: AB075751; BAB8861.1; -.
DR GO: GO:0016491; P:oxidoreductase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR001804; Isohd.
DR Pfam: PF00180; Isohd; 1.
SQ SEQUENCE 334 AA; 35922 MW; 82B018FED744FB49 CRC64;

Query Match       7.9%; Score 84.5; DB 2; Length 334;
Best Local Similarity 22.8%; Pred. No. 77;
Matches 50; Conservative 34; Mismatches 94; Indels 41; Gaps 7;

QY 10 VAFVAVSADPIHYDKITEEINKAIDAIAIEQESETIDPMKVPDHDHAKFERHVGVDFKG 69
Db      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
114 LVIVRENTGEGLYVEQRRYDLVAIADAVIDSKASERI-----G 151

QY 70 ELAMWNTEARGLKQMCKRGQGANV-----KGEIGVKAHLLIGVHHDIV---SMEYD 117
Db      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :

Db 152 RAALRIATGRPKRTLHIANKANVLPTQLGFDTVFKEAKDFELVNVDLIIVNCAMQLV 211

QY 118 LAYKLGDLPHTTHVISDI-QDFVAL--SLIESDEGNITWTSPEVRQFNVANVHGILSI 174
Db     :   :   :   :   :   :   :   :   :   :   :   :   :   :   :

Db 212 MRPERFDVIVTNNLGILLSDLAAGLVGGGLAPSGNIGTTFATFPVHGSAPDIACKGI 271

QY 175 LDPIFGVLSOLTAIFODTV-RKENTKVLAFAFKRELK 212
Db    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :

Db 272 ANPATAILS----AAAMLVDYLGEKAARKEKAVLDVLER 307

RESULT 75
QBAXK9 PRELIMINARY; PRt; 426 AA.

AC QBAXK9;
DC AC QBAXK9; DT 01-JUN-2003 (T-EMBLrel.. 24, Created)
DT 01-JUN-2003 (T-EMBLrel.. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel.. 25, Last annotation update)
DE Putative ABC oligo/dipeptide transport, ATP-binding protein.
GN BT4012.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
ON NCBI_TaxID=818;
RX [1].
RN SEQUENCE FROM N.A.
RC STRAIN-VPL-5482 / ATCC 29148;
RK MEDLINE=N22550826; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis." ;
Science 299:2074-2076(2003).
RW EMBL: AE016943; AAO79117.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005224; F:ATP binding; IEA.
DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003439; ABC transporter.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 426 AA; 46748 MW; 635D2514FF7FEF8 CRC64;

Query Match       7.9%; Score 84.5; DB 16; Length 426;
Best Local Similarity 23.4%; Pred. No. 1.le+02;
Matches 44; Conservative 36; Mismatches 67; Indels 41; Gaps 9;
```

```

RT host specificities."
RL Nature 417:459-463(2002)
DR EMBL, AF012026; AM38611.1; -
DR GO: 0016020; C:membrane, IEA.
DR GO: 0004871; F:signal transducer activity; IEA.
DR GO: 0006935; P:chemotaxis; IEA.
DR GO: 0007165; P:signal transduction; IEA.
DR InterPro: IPR004089; Chmtaxis_transd.
DR InterPro: IPR003660; HAMP.
DR InterPro: IPR004090; Me_Chemotaxis.
DR Pfam: PF00672; HAMP; 1.
DR Pfam: PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00885; HAMP; 2.
KW Complete Proteome.
SQ SEQUENCE 710 AA; 75422 MW; D84C7536ED0C3FD0 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 710;
Best Local Similarity 23.1%; Pred. No. 2.1e+02;
Matches 54; Conservative 31; Mismatches 80; Indels 69; Gaps 11

```

Db 609 K-----PYSVILFDEIEKAHPDVFNILLQVLDDGHLTDTKGRTVDFRNTVIIIMTSNVGAQ 663

```
QY 173 SILDPIFGVLSVLTAFQTVRKEMTKVLAPAKKE 209
Db 664 ELQDQRFAGGAGSESDYETVRKTKWMLKNSFRPE 700

RESULT 78
Q898W1 PRELIMINARY; PRT; 841 AA.
AC Q898W1
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE Protein translocase subunit secA.
GN SECA OR CTO00327.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / B88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA Decker J., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottechalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL PLoS Biol. 2:1316-1321 (2003).
DR EMBL; AEO15937; AAC34968.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006605; P:protein targeting; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000185; SecA.
DR InterPro; IPR004027; SEC_C motif.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF02810; SEC-C; 1.
DR Pfam; PF01043; SecA protein; 1.
DR PRINTS; PR00906; SECA.
DR PROSITE; PS01312; SECA; 1.
DR Complete proteome.
KW SEQUENCE 841 AA; 96627 MW; 183803C31D555FC4 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 841;
Best Local Similarity 22.8%; Pred. No. 2.6e+02;
Matches 60; Conservative 37; Mismatches 67; Indels 99; Gaps 16;

QY 23 DKITEINKA--IDDAIAAEQSEETIDPM---KVPD---HADKFRHVGIVDFKGLAM 73
Db 425 DEIVETHKGPVLVGTVSTKSELISEMLKRGVPHQVLNXYHKEAEIVSHAGEYGM 484

QY 74 ----RNTEARGLKQMGDANVKGEEGIVKAH--LLIGV--HDD-----110
Db 485 VTIATNAGRG-----TDKLEEEVVKAGGLKIIGTERHESRRIDNQLNGRSGRQG 535

QY 111 -----IVSMEVDL-----AYKGLDHPHTHVISD---IQDFVVALSLEISD 148
Db 536 DPGASRFVSLDDLMRIFGSDKLGIVEKLG-----LGDDAEIESQWVSNATENAQ 587

QY 149 ---EGNITMTSFVR---OFANVNH-----IGLSILDPIFGVLSDLVLTAI 189
Db 588 KQVEGN---NFDIRKTLIQYDDVINKQREIYKQSEVLEGADLKQIQEIMIRDVINSV 643

QY 190 FQDTRKEMTKVLAPAKRELEK 212
Db 644 VDHSID-----IEEFKEELDK 661

RESULT 79
Q89XK3 PRELIMINARY; PRT; 1429 AA.
AC Q89XK3
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE LD43558p.
GN RPII215 OR CG1554.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B. Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089555; AAL90293.1; -.
DR FlyBase; FBGN0003277; RPII215.
DR GO; GO:0005665; C:DNA-directed RNA polymerase II, core complex; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR006592; RNA_pola_N.
DR InterPro; IPR000684; RNA_poliI_repeat.
DR InterPro; IPR000722; RNA_poliA.
DR InterPro; IPR007066; RNA_poliRpb1_3.
DR InterPro; IPR007083; RNA_poliRpb1_4.
DR InterPro; IPR007081; RNA_poliRpb1_5.
DR InterPro; IPR007075; RNA_poliRpb1_6.
DR InterPro; IPR007073; RNA_poliRpb1_7.
DR Pfam; PF00623; RNA_poliRpb1_2; 1.
DR Pfam; PF04983; RNA_poliRpb1_3; 1.
DR Pfam; PF05000; RNA_poliRpb1_4; 1.
DR Pfam; PF04998; RNA_poliRpb1_5; 1.
DR Pfam; PF04992; RNA_poliRpb1_6; 1.
DR Pfam; PF04990; RNA_poliRpb1_7; 1.
DR Pfam; PF05001; RNA_poliRpb1_R; 22.
DR SMART; SMO0663; RPOLA_N; 1.
DR PROSITE; PS00115; RNA_POLI; REPEAT; 11.
DR SEQUENCE 1429 AA; 157524 MW; 540A045520CAEDA6 CRC64;

Query Match 7.9%; Score 84.5; DB 5; Length 1429;
Best Local Similarity 22.7%; Pred. No. 5.3e+02;
Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;

QY 17 ADPIHYDKITEINKADDAIAAEQSEETIDPMKVPDHA--DKERHVGIV-----DFKG 69
Db 228 ADPQTNIIEQAIKAKKDDVINVIQKARNLEPTPGTLRQTFENKVNILNDARDKTG 287

QY 70 ELAMRNI--EARGLKQMKRQ-----GDANVKGEE---GIVKAHLLIGVHD 109
Db 288 GSAKSLTEYNNLKVAMVSGSGKSNINISQIVACGQGVNVEGKEIPIYGRKRTLPFIKD 347

QY 110 D-----IVSMEVDLAYKGLDHPHTT---HVISDIQDFVVALSLEISDEGNI-----T 153
Db 348 DYGPESRGFVNSY-----LAGLTFSEFFYHAMGG-REGLIDTAVKTAETGYIQRLIKA 401

QY 154 MTSFEVRQFANVNHIGLSIL---DPIFGVLSD-----184
Db 402 MESVMVNDGTVRNSVGGLIQIRYGEDGLCGELVEFQNMPTVKLSNKSFEKPFKEDWSNE 461

QY 185 -VLTAIFQDTRKEMT 199
Db 462 RLMKKVFTDDVKEMT 477
```

RESULT 80
Q8KV19 PRELIMINARY; PRT; 186 AA.
AC Q8KV19; 186 AA.
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437230; AAM49491.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:Chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60/TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20429 MW; CEC0EAEC4A1A7FB CRC64;

Query Match 7.9%; Score 84; DB 2; Length 186;
Best Local Similarity 20.8%; Pred. NO. 39;
Matches 40; Conservative 35; Mismatches 63; Indels 54; Gaps 8;
QY 13 VAVSADPHYDKITEINKATDDAIAAEQSETIDPMKVPDHDADKPERHVGIVDFKELA 72
Db 16 VAGANPM-----DLKRGIDKAVAVETIGNAEKVDNDYDKIEQ-VATVSANDPA 67
QY 73 MRNTEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLYAYKLGDLHPTTH-- 130
Db 68 IGRLLAEAFKVKSGDGVITIEESK---SRETYIGV---VEGQDFDRGYISAYMAT 121
QY 131 -----VSDIQDF-----VVALSLSDSGNITMTSFEVRQFA 163
Db 122 ECQVDNPEYLLYDKKISNKEFLEILOPAESGRPLLVIAEDVDSEALTTL----- 172
QY 164 NVNHHI-GGLSI 174
Db 173 -VNRLRGLKI 183

RESULT 81
Q8KV15 PRELIMINARY; PRT; 186 AA.
AC Q8KV15; 186 AA.
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066(2002).

DR EMBL; AF437010; AAM49271.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:Chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60/TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20276 MW; 9B939C6DBCDF8F78 CRC64;
Query Match 7.9%; Score 84; DB 2; Length 186;
Best Local Similarity 23.0%; Pred. NO. 39;
Matches 43; Conservative 35; Mismatches 65; Indels 44; Gaps 10;
QY 13 VAVSADPHYDKITEINKATDDAIAAEQSETIDPMKVPDHDADKPERHVGIVDF 67
Db 16 VTAGANPM-----DLKRGIDKAVAVAVFIKHAEQVD-----DNYDKIEQ-VATVSA 62
QY 68 KGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYD-----LAYKLG 123
Db 63 NNDPEIGKLLADAMKVKSGDGVITIE-EKTMQTEL-----DLVEGQDFDRGYISAYMAT 116
QY 124 DLHP-----TTHVISDIDQDFVALSLSDSGNITMTSFEVRQFA--NVNHH 168
Db 117 DMKMEAVLDDPYLITDKKISNIDQLLFIQVSGARLLIAEDVDSEALTTLVNR 176
QY 169 I-GGLSI 174
Db 177 LRGLKI 183

RESULT 82
Q8DWU5 PRELIMINARY; PRT; 306 AA.
AC Q8DWU5; 306 AA.
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Transcriptional regulator, Cro/CI family.
GN SAG2118.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Kouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014287; AAN00976.1; -
DR TIGR; SAG2118; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF01381; HTH_3; 1.
DR SMART; SM00530; HTH_XRE; 1.
KW Complete proteome.
SQ SEQUENCE 306 AA; 35464 MW; 6B9E7E08B7BDEAFA CRC64;
Query Match 7.9%; Score 84; DB 16; Length 306;
Best Local Similarity 23.8%; Pred. NO. 76;
Matches 34; Conservative 25; Mismatches 62; Indels 22; Gaps 6;

```
QY 18 DPHYDKITEINKAI--DDAIAAEQSETIDPMKVPDHPADKFERHVGI--VDFKGLAMEN 75
Db 146 DDIFFMCELTKNISRQEAQTVLSQOEGLDTKIKELSNYPANQIYISYDFVGNL--EN 203
QY 76 IEARGHKKOMKQGDANVKEEGIVKAHLIGVHDDIVSMEYDLAYKLG-----DLHPT 128
Db 204 LSKDKTKQEVKE-----SYERLLKDSPOAG-----VEVSYSNLLGLFLINTLADSO 252
QY 129 THVISIDQFVVALSLEISDEGN 151
Db 253 RQLLSDMSQATVGMPSQFHEGH 275

RESULT 83
Q9WZQ7 PRELIMINARY; PRT; 314 AA.
AC Q9WZQ7;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein TM0800.
GN Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
HEidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001747; AAD35882.1; -
DR FNR; A72335; A72335.
DR HSSP; P12268; I530.
DR TIGR; TM0800; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR0041136; 2nprop dioxygen.
DR InterPro; IPR003009; FMN_enzyme.
DR Pfam; PF03060; NPD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 314 AA; 33673 MW; 3321EBCF7647D9B8 CRC64;

Query Match 7.9%; Score 84; DB 16; Length 314;
Best Local Similarity 23.5%; Pred. No. 78;
Matches 48; Conservative 32; Mismatches 76; Indels 48; Gaps 10;

QY 23 DKITEINKAIDDAIAA--IEQSETIDPMKVPDHPADKFERHVGI-----GIVDFKGLA 72
Db 122 DSLAMVERAGADAVIAEGNESGCHGEVTVTVLVNKSRSVNPVIAAGGADGRGMAA 181
QY 73 MRNIEARGLKMOKR-----QGDANVKEEGIVKAHLIGVHDDIVSMEYDLAYKGLDHP 127
Db 182 AFALGAQVQMGTRFVASVESDVHPYVKEKIVAA-----SIRDVTVT-----GAKLG--HP 230
QY 128 T-----THVISIDQFVVALSLEISD-----EGNITMTSFEVROFANVNHIGSL 172
Db 231 ARVLRTPFARKI GEMEPENPMQAEMLVGLSRRVVEGDLERGSFVWQSAGLIDSEI--- 287
QY 173 SILDPIFGVLSVLTATFQDTEVK 196
Db 288 ---KPVKQIIEDILKE--FKETVEK 307
```

RESULT 84

```
Q99X76 PRELIMINARY; PRT; 392 AA.
AC Q99X76;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phosphotransferase (EC 5.4.2.7).
GN DRM OR SAV0139 OR SA0134 OR MW0113.
OS Staphylococcus aureus (strain Mu50 / ATCC-700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RL aureus.";
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yanamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AP003358; BAB56301.1; -
DR EMBL; AP003129; BAB41354.1; -
DR EMBL; AP004822; BAB93978.1; -
DR FNR; G89774; G89774.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0008973; F:phosphotransferase activity; IEA.
DR InterPro; IPR006124; Metalloenzyme.
DR Pfam; PF01676; Metalloenzyme; 1.
KW Isomerase; Complete proteome.
SQ SEQUENCE 392 AA; 43795 MW; 6012D71CD9BDB549 CRC64;

Query Match 7.9%; Score 84; DB 16; Length 392;
Best Local Similarity 24.1%; Pred. No. 1e+02;
Matches 45; Conservative 35; Mismatches 85; Indels 22; Gaps 10;

QY 14 AVSADPHYDKITE--EINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDFKGLA 72
Db 65 AVEQPEAVTKLSASVCKDTMTCHWEIMGLNITQPKVYPNGPPELIQOIEWTKKV 124
QY 73 MRNIEARGLKMOKRQGDANVKEEGIVKAH-----LLIGVHDDIVSME--YDLAYKGLD- 125
Db 125 VANKPASQTQIIDEWGEHOMKTGLIVVTSADPVLQIAAHEDIPLBELYDICEKVELT 184
QY 126 HPTHTVISIDQFVVALSLEISDEGNITMTS-----FEVROFA-NVNHIGSLDPI 178
Db 185 KDPKYLIGRI-----IARPY-VGEFGNFTRTSNRHDYALKPFGKTVLHDKGGYDVI--A 237
QY 179 FGVLSDV 185
Db 238 IGKINDI 244

RESULT 85
Q9X9I6 PRELIMINARY; PRT; 600 AA.
ID Q9X9I6
```


AC Q9X9I6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Inner membrane ABC-transporter, Irp7.
 GN IRP7.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99426823; PubMed=10496905;
 RA Rakin A., Noelting C., Schubert S., Heesemann J.;
 RT "Common and specific characteristics of the high-pathogenicity island
 of Yersinia enterocolitica.";
 RL Infect. Immun. 67:5265-5274(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99426823; PubMed=10496905;
 RA Rakin A., Schubert S., Gullvout I., Carniel E., Heesemann J.;
 RT "Local hopping of IS3 elements into the A-T-rich part of the high-
 pathogenicity island in Yersinia enterocolitica 1B, O:8.";
 RL FEMS Microbiol. Lett. 182:2255-2294(2000).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AJ132668; CAB46572.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; P:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC TM transp.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00664; ABC membrane; 1.
 DR Pfam; PF00005; ABC trans; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 600 AA; 66364 MW; F6C6411A26737E4 CRC64;
 Query Match 7.9%; Score 84; DB 2; Length 600;
 Best Local Similarity 22.2%; Pred. No. 1.8e+02;
 Matches 43; Conservative 25; Mismatches 66; Indels 60; Gaps 6;
 QY 6 LIAAVAFVAVSADP-----IHYDKITEINKAIDDAIAAIOSETIDPMKVPDHDADKPERH 61
 DB 287 LIAAVAMIMRFAEPFAMFISTVSVE-----LIASALQRIEQFVAIAPLPVAAQSEMPERY 342
 QY 62 VGIVDFKGEALARNIENARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSVEYDLAYK 121
 DB 343 -----DIRFDNVSR-----
 QY 122 LGDLHPHTVISDIQDFVALSLRISDEGNITMTSPFVROFANVFN---HIGGLSLIDPI 178
 DB 355 EGDGHALNHLSTLTPAASMSALVAGSAGKTTVTKLMR-YADPQQQISIGGVDIRLRT 413
 QY 179 FGVLSDVLTAIFQD 192
 DB 414 PEQLNLSISVVFQD 427
 RESULT 86
 Q92375 PRELIMINARY; PRT; 600 AA.
 AC Q92375;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE YETQ protein (Inner membrane ABC-transporter YbtQ) (Permease and
 DE ATP-binding protein of yersiniabactin-iron ABC transporter).
 GN YBTQ OR YP01914 OR Y2396.

OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=6/69;
 RA Buchrieser C., Rueniok C., Couve E., Frangeul L., Billault A.,
 RA Kunst F., Carniel E., Glaser P.;
 RT "DNA sequence of the 102 kbases unstable region of Yersinia pestis.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=KIM6;
 RA MEDLINE=99035519; PubMed=9818149;
 RA Gehring A.M., DeMoll E., Fetherston J.D., Mori I., Mayhew G.F.,
 RA Blattner F.R., Walsh C.T., Perry R.D.;
 RT "Iron acquisition in plague: modular logic in enzymatic biogenesis of
 yersiniabactin by Yersinia pestis.";
 RL Chem. Biol. 5:573-586(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holroyd S., Jageis K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmons M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AL031866; CAA21387.1; -.
 DR EMBL; AF091251; AAC69584.1; -.
 DR EMBL; AJ414150; CAC90730.1; -.
 DR EMBL; AE013842; AAM85954.1; -.
 DR FIR; AF0233; AF0233.
 DR FIR; TJ7436; TJ7436.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001140; ABC TM transp.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00664; ABC membrane; 1.
 DR Pfam; PF00005; ABC trans; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
 KW ATP-binding; Transport; Complete proteome.
 SQ SEQUENCE 600 AA; 66464 MW; 759638E8250E559C CRC64;
 Query Match 7.9%; Score 84; DB 16; Length 600;
 Best Local Similarity 21.6%; Pred. No. 1.8e+02;
 Matches 42; Conservative 27; Mismatches 65; Indels 60; Gaps 6;
 QY 6 LIAAVAFVAVSADP-----IHYDKITEINKAIDDAIAAIOSETIDPMKVPDHDADKPERH 61
 DB 287 LIAAVAMIMRFAEPFAMFISTVSVE-----LIASALQRIEQFVAIAPLPVAAQSEMPERY 342

```
QY 62 VGIIVDFKGLAMRNIEARGLKQKMQGDANVKGEEGIVKAHLIIHVHDDIVSMEYDLAYK 121
Db 343 -----DIRFNVSYR-----YE 354
QY 122 LGDLHPTTHVISDIQFVVALSLEISDEGNITWTSFEVQFANVNV---HIGLSILDP 178
Db 355 EGDGHALNHVSLTPFAASMSALVGASGAGKTTVTKLLMR-YADPQOGQISIGVDIRRLT 413
QY 179 FGVLSVLTALFOD 192
Db 414 PEOLNLSISVVD 427

RESULT 87
Q9CHZ3 PRELIMINARY; PRT; 775 AA.
AC Q9CHZ3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Cation-transporting ATPase.
GN YFGQ OR LL0575.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
CX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis sep. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AEO06291; AAK04673.1; -.
DR PIR; G86696; G86696.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0006812; P:cation transport; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001757; ATPase_E1-E2.
DR InterPro; IPR008250; E1-E2_ATPase_reg.
DR InterPro; IPR005834; Hydrolyase.
DR Pfam; PF00122; E1-E2_ATPase; 1.
DR PRINTS; PR00119; CATATPASE.
DR TIGRFAMs; TIGR01494; ATPase_P-type; 4.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
KW Complete proteome.
SQ SEQUENCE 775 AA; 85467 MW; 84D5E4E49A833BE CRC64;

Query Match 7.9%; Score 84; DB 16; Length 775;
Best Local Similarity 19.8%; Pred. No. 2.6e+02;
Matches 50; Conservative 3; Mismatches 103; Indels 62; Gaps 6;

QY 22 YDKITEINKAIDATAAIEQSEITDPMKVPDHDKFEHVGIVDFKGL-----AMEN 75
Db 397 YEKVSINHLLEGFRVLVAGTKE--KIYDQNLNGALGVGVLIANPIRENAKSTFNY 454
QY 76 IEARGLKQKMQGD-----ANVKGEEGIVKAHLI----- 104
Db 455 FAEGGVNKKVIGSDNPQTVSAVAKRAGITGAERFIDANLLKTKEDLDQAVESYTVFGRYT 514
QY 105 -----IGVHDDIVSMEYD-----LAYKGLDHPHPTTHVISDIQDFVVALSLEISDE 149
Db 515 PDQKRRLVQALKRKHDTVAMTGVGVNDILAMKSDCSIAMASGSDAATQVAVVLLDSDF 574
QY 150 GNTITWTSFEVQFANVNVHIGLSILDPVGVLSVLTALFODTVKEMTK-----V 201
Db 575 GHMTQVTEGRRVNVNQRSAIFLVKVLFIILAIISALFVFTYPLQASQLSLISLFTI 634
```

```
QY 202 LAPAFKRELEKN 213
Db 635 GIPGFLLSLEEN 646

RESULT 88
Q93NX9 PRELIMINARY; PRT; 9510 AA.
AC Q93NX9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE AmphI.
GN AMPHI.
OS Streptomyces nodosus.
OC Bacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., O'Liynyk M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
RT nodosus."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AF357202; AAK73501.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0004334; F:lacyl-carrier protein] S-malonyltransferase. . .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; AC trans.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR004410; FAD.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR001005; Myb_DNA_Binding.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp bind.
DR Pfam; PF00698; Acyl transf; 6.
DR Pfam; PF00106; adh short; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fadD; 6.
DR PROSITE; PS00075; ACP DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00037; MYB 1; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 6.
KW Oxidoreductase; Phosphopantetheine; Transferase.
SQ SEQUENCE 9510 AA; 992817 MW; 1ABE003A15D478CF CRC64;

Query Match 7.9%; Score 84; DB 2; Length 9510;
Best Local Similarity 24.5%; Pred. No. 7.1e+03;
Matches 53; Conservative 39; Mismatches 80; Indels 44; Gaps 10;

QY 3 KELLIAAVFVAVSADPHYD-----KIT--EEINK-----AIDDAIAAIEQSE 44
Db 582 RFPVFAALDAALDAFTPLHDPVLRKVLWGSDADRLDRTEYAQPALFAVEVALYLLSEF 641
QY 45 TIDPMKVDPDA--DKFERHVGIVDFKGLAMRNIEARG-LQMKRQGDANVKGEEGIVKA 101
Db 642 EVKPDHLGSHVGIEAAAHVAGV-PSLDDAATLVAARGRLQALPEGGMVAVQASEDEV 700
QY 102 HLLIGVHDDIVSMEYDILAYKGLDHPHPTTHVISDIQDFVVALSLEISDEGNIT----- 153
Db 701 AFLLAGHEDLVs-----LAAVNG---PSAVVLGSDETTVELAARLAADGRKTSRLRVSHA 753
QY 154 ---WTSFEVQFANVNVHIGLSILDPVGVLSV 185
Db 754 FHSPLMAPMLDEFNRVVE---GLTHSPLLPVVSDV 786
```

```

RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Urwin L., Whitehead S., Barrell B.G., Maskell D.J., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43727.1; -.
KW Complete proteome.
SQ SEQUENCE 266 AA; 28658 MW; 7207543ADBFF8F6F CRC64;

Query Match 7.8%; Score 83.5; DB 16; Length 266;
Best Local Similarity 24.4%; Pred. No. 69;
Matches 53; Conservative 32; Mismatches 75; Indels 57; Gaps 12;

QY 3 KFLIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-AIAAIEQSETID---PMK 50
Db 17 RYLATAVVALAVAAALYARQDEFEHDEV---VLLCLDDQAMSTVTRTGSVDGILPAC 73
QY 51 VPDHAKFERHVGIVDFKGLAMRNIEAR-GLKQKQGDANVKGEGIVKAHL----- 104
Db 74 QPELAEVLGEDFGAVD---EPTLRVAVLATVAANFAEYSSSAVACEAIKSOHLNCGNL 130
QY 105 -----IGVHDDIVSMYDLYAKLGD---LHPTTHVISDIQDFVVAL 142
Db 131 VFLMGVLYGLKEKGIPIGVGGAGVGHQAQMTKSGNNRLLDPTTGVVAAI-TFNKLM 189
QY 143 SLEISDEGNI-----TMTSFEVRQFANVNNHIG 171
Db 190 SGQPVPEQVRVFRVKATIDSFDRVYAVVG--GG 224

RESULT 91
Q93R43 PRELIMINARY; PRT; 311 AA.
AC Q93R43;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
GN HSP60.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=630;
RN [1]_TaxID=630;
RP SEQUENCE FROM N.A.
RA Yang R.P.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Du Z.M.;
RT "Bacillus and Versinia HSP60 gene partial sequence.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064591; BAB61892.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1 311
FT NON_TER 311 311
SQ SEQUENCE 311 AA; 33032 MW; 34429B1392166D86 CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 311;
Best Local Similarity 18.1%; Pred. No. 85;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VASADPHYDKITEINKAIDDA-IAAIEQSETIDPMKVPDHPADKFERHVGIVDFKGL 71
Db 16 VAAGNPNM-----DLKRGIDKAVIAAVES---LKKLSVPCSDSKAIAQVGTISANDS 65
QY 72 AMRNIARGLKQMKRQGDANVKGEGIVKAHLIGVHDD---IVSMYDLYAKLGD 120

RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Urwin L., Whitehead S., Barrell B.G., Maskell D.J., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43727.1; -.
KW Complete proteome.
SQ SEQUENCE 266 AA; 28658 MW; 7207543ADBFF8F6F CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 186;
Best Local Similarity 24.1%; Pred. No. 43;
Matches 41; Conservative 25; Mismatches 65; Indels 39; Gaps 8;

QY 13 VASADPHYDKITEINKAIDDAIAIEQSETIDPMKVPDHPADKFERHVGIVDFKGLA 72
Db 16 VTAGASPI-----DLKRGIDKAVAAVAVELKAHSHEVGDDYKVEQ-VGTVAANDAT 67
QY 73 MRNIEARGLKQMKRQGDANV---KGEIGVKAHLIGVHDDIVSMYDLYAKLGD.LHPTT 129
Db 68 IGKLIADAMSKVKDGVITVEEAKGTETVKV-----VEGMQFDRGY----ISPYF 114
QY 130 HVISDIQDFVVALSLEIS-DEGNTMTSFEVRQFANVNNHIGLSILDPI 178
Db 115 MTNSD-----KMEASLSDSPYVLTIDTKKISTMKDV-----LPFILEPI 150

RESULT 90
Q7VTQ1 PRELIMINARY; PRT; 266 AA.
AC Q7VTQ1;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative exported protein.
GN BP3464
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-Tohama I // ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.N., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
```

```

Db 66 TVGELIAQAMEKV-----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 116
QY 121 -----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 159
Db 117 PETGSIELESFILLADKKISNIREMLPVLEAVAKAGKPLIIAEDVEGEALATL----- 171
QY 160 RQFANVNNHIGGL-----SILDPIFGVLSVDLTAIFQDVTVRKEMTKVLAPAFKRELEK 212
Db 172 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEIGLELEK 220

RESULT 92
Q93R44
ID Q93R44 PRELIMINARY; PRT; 311 AA.
AC Q93R44
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
GN HSP60.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
SEQUENCE FROM N.A.
RA Yang R.F.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA Du Z.M.;
RT "Bacillus and Yersinia HSP60 gene partial sequence.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064592; BAB61893.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:Chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 311
FT NON_TER 311
SQ SEQUENCE 311 AA; 33002 MW; C1428E0692166D9C CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 311;
Best Local Similarity 18.1%; Pred. No. 85;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSETIDPMKVPDHPADKFERHVGIVDFKGE 71
Db 16 VAGMNP-----DLKRGIDKAVIAAEE---LKKLSVPCSDSKAIAQVGTISANS 65
QY 72 AMRNIARGLKQMKROGDANVKGEIGVKAHLIIGVHDD---IVSMEYDLAY----- 120
Db 66 TVGELIAQAMEKV-----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 116
QY 121 -----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 159
Db 117 PETGSIELESFILLADKKISNIREMLPVLEAVAKAGKPLIIAEDVEGEALATL----- 171
QY 160 RQFANVNNHIGGL-----SILDPIFGVLSVDLTAIFQDVTVRKEMTKVLAPAFKRELEK 212
Db 172 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEIGLELEK 220

RESULT 94
Q93R44
ID Q93R44 PRELIMINARY; PRT; 329 AA.
AC Q93R44
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE groEL-like protein (60 kDa chaperonin) (Protein Cpn60) (groEL protein)
DE (Fragment).
GN GROEL.
OS Enterobacteriaceae sp. JN965.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae.
OX NCBI_TaxID=146477;
RN [1]
SEQUENCE FROM N.A.
RA Rascoc J.E.; Berg M.; Bruton B.; Pair S.; Mitchell F.; Melcher U.;
RA Fletcher J.;
RT "Characterization of Serratia isolates associated with cucurbit yellow
vine disease.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
```

```

Db 66 TVGELIAQAMEKV-----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 116
QY 121 -----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 159
Db 117 PETGSIELESFILLADKKISNIREMLPVLEAVAKAGKPLIIAEDVEGEALATL----- 171
QY 160 RQFANVNNHIGGL-----SILDPIFGVLSVDLTAIFQDVTVRKEMTKVLAPAFKRELEK 212
Db 172 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEIGLELEK 220

RESULT 92
Q93R44
ID Q93R44 PRELIMINARY; PRT; 311 AA.
AC Q93R44
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
GN HSP60.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
SEQUENCE FROM N.A.
RA Yang R.F.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA Du Z.M.;
RT "Bacillus and Yersinia HSP60 gene partial sequence.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064590; BAB61891.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:Chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 311
FT NON_TER 311
SQ SEQUENCE 311 AA; 33002 MW; C1428E0692166D9C CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 311;
Best Local Similarity 18.1%; Pred. No. 85;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSETIDPMKVPDHPADKFERHVGIVDFKGE 71
Db 16 VAGMNP-----DLKRGIDKAVIAAEE---LKKLSVPCSDSKAIAQVGTISANS 65
QY 72 AMRNIARGLKQMKROGDANVKGEIGVKAHLIIGVHDD---IVSMEYDLAY----- 120
Db 66 TVGELIAQAMEKV-----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 116
QY 121 -----KLGDLPHPTTHVTSIDIQDFVVALSLEISDEGNITMTSFEV 159
Db 117 PETGSIELESFILLADKKISNIREMLPVLEAVAKAGKPLIIAEDVEGEALATL----- 171
QY 160 RQFANVNNHIGGL-----SILDPIFGVLSVDLTAIFQDVTVRKEMTKVLAPAFKRELEK 212
Db 172 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEIGLELEK 220

RESULT 93
Q93R44
ID Q93R44 PRELIMINARY; PRT; 311 AA.
AC Q93R44
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
```

EMBL; AJ302157; CAC19389.1; -
 DR HSP; P06139; 10RL.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003754; F:chaperone activity; IEA.
 DR InterPro; IPR001844; Chaptnin_Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60 TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 KW ATP-binding; Chaperone.
 FT NON_TER 329 329
 SQ SEQUENCE 329 AA; 34639 MW; F6C7177B5E8B15 CRC64;
 Query Match 7.8%; Score 83.5; DB 2; Length 329;
 Best Local Similarity 18.1%; Pred. No. 92;
 Matches 43; Conservative 44; Mismatches 81; Indels 69; Gaps 9;
 QY 13 VASADPHYDKITEINKAIDDAI-AAIQSEITIDPKVPDHDADKPERHVGIVDFKGL 71
 DB 107 VAAGNNPW-----DLKRGIDKAVTAABE---LKALSVPSCDSKAIAQVGISANSDE 156
 QY 72 AMRIEARGLKOMKQGDANVKGEGIVKAHLIGVHDD---IVSMEYDLAY----- 120
 DB 157 TVGKLIAEMKV-----GREGVITVEGTLQDELVDVVGQFDRGYLSPYFINK 207
 QY 121 -----KLGDLHTTHVDSIQDFVVALSLEISDEGNITWTSFEV 159
 DB 208 PETGSVELESFILLADKKISNIREMLFVLEAVAKAGKPLVIAEDVEGALATL----- 262
 QY 160 RQFANVNHGGL---SILDPIGVLSDLTAIFQDTVRKMTKVLAPAFKRELEX 212
 DB 263 -----VNTMRGIVKVAVKAPFG---DRKAMLDIATITGTVIASEIGHCELEX 311
 RESULT 95
 Q8K3U8 PRELIMINARY; PRT; 402 AA.
 AC Q8K3U8
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE FK506 binding protein 4 (Fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RA Desai B.J., McKinney K.Q., Meyer M.H., Bahrani-Mostafavi Z.,
 RA Meyer R.A. Jr.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF531427; AAM95632.1; -
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR001179; FKBP_PPIase.
 DR InterPro; IPR001440; TPR.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF00254; FKBP; 2.
 DR Pfam; PF00315; TPR; 3.
 DR SMART; SM00028; TPR; 3.
 DR PROSITE; PS00454; FKBP_PPIASE_2; 2.
 DR PROSITE; PS00059; FKBP_PPIASE_3; 2.
 FT NON_TER 1
 SQ SEQUENCE 402 AA; 45486 MW; 82D74DB9643D1B51 CRC64;
 Query Match 7.8%; Score 83.5; DB 11; Length 402;
 Best Local Similarity 22.1%; Pred. No. 1.2e+02;
 Matches 27; Conservative 30; Mismatches 48; Indels 17; Gaps 3;
 QY 18 DPHYDKITEINKAIDDAIAIEQSETI-----DPMKVPDHDAD-KFERHV 62
 DB 19 DKFSPDLGKGEVKAWDIAVATMKVGVCHITCKPEYAGSGSPKPPNATLVFE--V 76

QY 63 GIVDFKGLAMRNIEARGLKOMKQGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKL 122
 DB 77 ELFEFKGEDLTDEDDGGIIRTRGEGYARFNDGAMVVALEGGYNDRLFDQRELCEV 136
 QY 123 GD 124
 DB 137 GE 138
 RESULT 96
 Q8NUB6 PRELIMINARY; PRT; 412 AA.
 AC Q8NUB6
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 5'-nucleotidase/2',3'-cyclic phosphodiesterase and related
 DE esterases.
 GN CGL0023.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 ON NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF005274; BAB97416.1; -
 DR GO; GO:0016787; P:hydrolase activity; IEA.
 DR GO; GO:0009166; P:nucleotide catabolism; IEA.
 DR InterPro; IPR008334; 5'-Nucleotidase C.
 DR InterPro; IPR006179; 5'-Nucleotidase.
 DR InterPro; IPR001064; Cystallin.
 DR Pfam; PF02872; 5_nucleotidaseC; 1.
 DR PRINTS; PR01607; APYRASEFAMILY.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 KW Complete proteome.
 SQ SEQUENCE 412 AA; 43019 MW; C48CA12D657A5B02 CRC64;
 Query Match 7.8%; Score 83.5; DB 16; Length 412;
 Best Local Similarity 31.5%; Pred. No. 1.2e+02;
 Matches 34; Conservative 16; Mismatches 41; Indels 17; Gaps 6;
 QY 54 HADKFERHVGIVDFKGLAMRNIEARGLKOMKQGDANVKGEGIVKAHLIGVHDDIVS 113
 DB 134 YAEAFE---IQFSGEDSFVTLKGDFKDALEQ-----QWEEGSAFPVAALGVSDN-VS 183
 QY 114 MEYDLAYKGLDLPHTHTVISDI-----ODFVVALSLEISDEGNITWTS 156
 DB 184 YTYDINRPIGD-RVTSVTIDTLPDLPDPERDYVVAASLYL-QSGNEGTA 229
 RESULT 97
 Q97K37 PRELIMINARY; PRT; 474 AA.
 AC Q97K37
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Beta-glucosidase family protein.
 GN CAC1084
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 ON NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RA MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,

[illegible]

Mon Aug 9 08:14:05 2004

```

DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 20213 MW; E5919F6D83F16A0E CRC64;

Query Match 7.8%; Score 83; DB 2; Length 186;
Best Local Similarity 23.3%; Pred.No. 47;
Matches 27; Conservative 26; Mismatches 49; Indels 14; Gaps 4;

QY 5 LLIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEQSETIDPMKVPDRAKFERHVG 64
Db 8 ILTEGMKVAAAGAPL-----DVKRGIDKAVAKVVEIKSQAEVGDYDKIEQ-VGT 59

QY 65 VDFKGEIARNIIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLY 120
Db 60 VSANNDPVIGKLIADAMRKVKSGDVITIEAKG---TDTTIGV---VEGMQFDRGY 109

```

Search completed: August 6, 2004, 16:01:44
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 16:02:17 ; Search time 54 Seconds
(without alignments)
1114.493 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213

Sequence: 1 MKKLLIAAVFAVVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A_Geneseq_29Jan04:**

1: Geneseqp1980s:**

2: Geneseqp1990s:**

3: Geneseqp2000s:**

4: Geneseqp2001s:**

5: Geneseqp2002s:**

6: Geneseqp2003as:**

7: Geneseqp2003bs:**

8: Geneseqp2004s:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	213	2	AAR60576 House dus
2	213	100.0	213	2	AAY25592 D. farina
3	213	100.0	213	7	ADC3482 House dus
4	25	11.7	215	2	AAR60575 House dus
5	25	11.7	215	2	AAY25586 D. pteron
6	25	11.7	215	7	ADC34836 House dus
7	8	3.8	87	5	ABP27647 Streptoco
8	8	3.8	321	3	AAG31332 Arabidops
9	8	3.8	342	3	AAG31331 Arabidops
10	8	3.8	349	3	AAG31330 Arabidops
11	8	3.8	412	2	AAR41823 Methyloba
12	8	3.8	484	6	ABU23355 Protein e
13	8	3.8	580	2	AAR68402 FeLV-A ga
14	8	3.8	611	4	AAM50132 Feline le
15	8	3.8	783	5	ABBS2052 Herbicida
16	8	3.8	1784	2	AAR05898 Gene prod
17	8	3.8	1784	2	AAR34427 FeLV F6A
18	7	3.3	73	3	AAB51943 Human sec
19	7	3.3	85	6	ABR41325 Human DIT
20	7	3.3	87	4	AAU37687 Streptoco
21	7	3.3	87	5	ABP27648 Streptoco
22	7	3.3	87	6	ABU01601 S. pneumo
23	7	3.3	87	6	ABU46065 Protein e
24	7	3.3	87	6	ABU44650 Protein e
25	7	3.3	87	6	ABU46703 Protein e

26	7	3.3	133	3	AAG01133	Aag01133 Human sec
27	7	3.3	139	4	AAU30907	Aau30907 Novel hum
28	7	3.3	148	6	ADA33025	Ada33025 Acinetoba
29	7	3.3	152	4	AAU30909	Aau30909 Novel hum
30	7	3.3	155	3	AAG61764	Aag61764 Arabidops
31	7	3.3	178	5	ABP26815	Abp26815 Streptoco
32	7	3.3	178	7	ADD71115	Add71115 Human int
33	7	3.3	181	3	AAG59755	Aag59755 Arabidops
34	7	3.3	183	3	AAG59754	Aag59754 Arabidops
35	7	3.3	183	3	AAG55907	Aag55907 Arabidops
36	7	3.3	185	3	AAG55324	Aag55324 Arabidops
37	7	3.3	190	5	ABP69706	Abp69706 Human pol
38	7	3.3	192	7	ADA24313	Ada24313 Bacillus
39	7	3.3	193	5	ABB06132	Abb06132 Human NS
40	7	3.3	204	2	AAR54114	Aar54114 GDI D4 pr
41	7	3.3	205	3	AAG59753	Aag59753 Arabidops
42	7	3.3	214	3	AAG55323	Aag55323 Arabidops
43	7	3.3	215	3	AAG55906	Aag55906 Arabidops
44	7	3.3	267	7	ADB64094	Adb64094 Human pro
45	7	3.3	295	7	ADB85214	Adb85214 Rat chima
46	7	3.3	300	3	AAB18216	Aab18216 Plasmodiu
47	7	3.3	343	6	ADB09304	Adb09304 Alloiooc
48	7	3.3	354	2	AAR33439	Aar33439 Ornithine
49	7	3.3	357	2	AAR33440	Aar33440 Ornithine
50	7	3.3	368	6	ABU35415	Abu35415 Protein e
51	7	3.3	372	6	ABU40439	Abu40439 Protein e
52	7	3.3	383	5	ABP69823	Abp69823 Human pol
53	7	3.3	425	6	ADB09306	Adb09306 Alloiooc
54	7	3.3	430	6	ABU23589	Abu23589 Protein e
55	7	3.3	434	5	ABBS4464	Abbs4464 Lactococc
56	7	3.3	453	6	ABU24395	Abu24395 Protein e
57	7	3.3	460	4	ABBS3535	Abbs3535 Drosophil
58	7	3.3	466	6	ABRS58660	Abrs58660 Human can
59	7	3.3	468	6	ABU25000	Abu25000 Protein e
60	7	3.3	468	7	ADE59897	Ade59897 Human pro
61	7	3.3	474	5	ABBS2783	Abbs2783 Herbicida
62	7	3.3	476	6	ADB09308	Adb09308 Alloiooc
63	7	3.3	498	6	ABP71433	Abp71433 Y. pestis
64	7	3.3	511	4	ABBS2629	Abbs2629 Drosophil
65	7	3.3	521	6	ADB09310	Adb09310 Alloiooc
66	7	3.3	535	6	ABJ39024	Abj39024 Mouse cla
67	7	3.3	535	7	ADC17327	Adc17327 Mouse IFN
68	7	3.3	581	4	ABGS5418	Abgs5418 Novel hum
69	7	3.3	581	5	ABG97496	Abg97496 Human NOV
70	7	3.3	625	6	ABU20412	Abu20412 Protein e
71	7	3.3	657	7	ADB80157	Adb80157 Mycobacte
72	7	3.3	752	2	AAR38153	Aar38153 Acetobact
73	7	3.3	791	5	ABBS9713	Abbs9713 Herbicida
74	7	3.3	887	6	ADB11976	Adb11976 Alloiooc
75	7	3.3	1001	6	ABU21679	Abu21679 Protein e
76	7	3.3	1027	7	ADC01332	Adc01332 Snteroha
77	7	3.3	1194	6	ABM15822	Abm15822 Mycobacte
78	7	3.3	1400	7	ADC01365	Adc01365 Enteroha
79	7	3.3	2000	6	ABRS2951	Abrs2951 Protein s
80	7	3.3	2836	4	ABBS6719	Abbs6719 Drosophil
81	6	2.8	8	4	ABBS23790	Abbs23790 Voltage s
82	6	2.8	8	5	ABG70880	Abg70880 R. rhodo
83	6	2.8	10	2	AAR70305	Aar70305 Laminin (
84	6	2.8	10	5	ABG67586	Abg67586 Human ADP
85	6	2.8	10	6	ABUS8121	Abu58121 Laminin c
86	6	2.8	10	6	ADA23705	Ada23705 Alzheimer
87	6	2.8	11	2	AAR69442	Aar69442 Rat phosph
88	6	2.8	15	2	AAW73718	Aaw73718 M. tuberc
89	6	2.8	15	2	AAW73716	Aaw73716 M. tuberc
90	6	2.8	15	2	AAW73826	Aaw73826 M. tuberc
91	6	2.8	15	2	AAW73828	Aaw73828 M. tuberc
92	6	2.8	15	4	AAU08183	Aau08183 Mycobacte
93	6	2.8	15	4	AAU08181	Aau08181 Mycobacte
94	6	2.8	19	2	AAR25844	Aar25844 Laminin-p
95	6	2.8	19	2	AAR44034	Aar44034 Laminin-d
96	6	2.8	19	2	AAR92741	Aar92741 IKVAV-con
97	6	2.8	19	2	AAY22410	Aay22410 Laminin A
98	6	2.8	19	2	AAY31007	Aay31007 Non-cross

99 6 2.8 19 5 AA48193 Aam48193 Peptide #
100 6 2.8 19 6 AB90762 Abu90762 Peptide #

ALIGNMENTS

RESULT 1
AAR60576
ID AAR60576 standard; protein, 213 AA.
XX
AC AAR60576;
XX
DT 25-MAR-2003 (revised)
DT 01-APR-1995 (first entry)
XX
DE House dust mite allergen DerfVII.
XX
KW DerfVII allergen; antiallergic; allergy diagnosis.
XX
OS Dermatophagoides farinae.
XX
PN WO9420614-A1.
XX
PD 15-SEP-1994.
XX
PF 11-MAR-1994; 94WO-AU000117.
XX
PR 12-MAR-1993; 93US-00031141.
PR 22-JUN-1993; 93US-00081540.
XX
PA (CHIL-) INST CHILD HEALTH RES.
XX
PI Thomas WR, Chua K;
XX
DR WPI; 1994-303021/37.
DR N-PSDB; AAQ71401.
XX
PT New nucleic acid encoding specific dust mite allergens - and related
PT vectors, transformed cells, peptides and antibodies, useful for
PT desensitisation and diagnosis.
XX
PS Claim 5; Page 40-41; 67pp; English.
XX
CC DerfVII antigen is useful as antiallergic reagent for treating
CC sensitivity to house dust mite allergens. (updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 213 AA;
Query Match 100.0%; Score 213; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e-203;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMKFLIIAAVAFVAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDADKFER 60
DB 1 MMKFLIIAAVAFVAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDADKFER 60
QY 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120
QY 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNHIGLSILDPFG 180
DB 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNHIGLSILDPFG 180
QY 181 VLSVLTATFQDVTVRKEMTKVLAPAFKLEKN 213
DB 181 VLSVLTATFQDVTVRKEMTKVLAPAFKLEKN 213

RESULT 2
AAY25592

ID AAY25592 standard; protein, 213 AA.
XX
AC AAY25592;
XX
DT 30-SEP-1999 (first entry)
XX
DE D. farinae allergen Der f 7 protein fragment.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Dermatophagoides farinae.
XX
PN WO9934826-A1.
XX
PD 15-JUL-1999.
XX
PF 11-JAN-1999; 99WO-CB000080.
XX
PR 09-JAN-1998; 98GB-00000445.
PR 21-SEP-1998; 98GB-00020474.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Larche M, Kay AB;
XX
DR WPI; 1999-458255/38.
XX
PT Desensitizing patients to polypeptide allergens.
XX
PS Example 6; Page 52; 117pp; English.
XX
CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitizing patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents the Dermatophagoides farinae allergen Der f 7
XX
SQ Sequence 213 AA;
Query Match 100.0%; Score 213; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e-203;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMKFLIIAAVAFVAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDADKFER 60
DB 1 MMKFLIIAAVAFVAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDADKFER 60
QY 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120
QY 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNHIGLSILDPFG 180
DB 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNHIGLSILDPFG 180
QY 181 VLSVLTATFQDVTVRKEMTKVLAPAFKLEKN 213

```
Db 181 VLSDLVLTAFQDITVRKEMTKVLAPAPKRELEKN 213

RESULT 3
ADC34842
ID ADC34842 standard; protein; 213 AA.
XX AC ADC34842;
XX AC ADC34842;
DT 18-DEC-2003 (first entry)
XX DE
XX DE House dust mite allergen Der f 7.
XX KW house dust mite; allergen; antigen; hyporesponsive; desensitisation;
XX KW immunomodulator; gene therapy.
XX OS Dermatophagoides farinae.
XX PN WO2003047618-A2.
XX PD 12-JUN-2003.
XX PF 05-DEC-2002; 2002WO-GB005548.
XX PR 05-DEC-2001; 2001US-0338385P.
XX PA (CIRC-) CIRCASSIA LTD.
XX PI Larche M, Ledger PW;
XX WPI; 2003-523267/49.
XX Desensitizing an individual to a selected polypeptide antigen comprises
PT administering a composition containing polypeptide antigens in an amount
PT that generates a state of hyporesponsiveness to the antigen to allow
PT desensitization.
XX PS Disclosure; Page 23; 57pp; English.
XX CC The invention relates to a novel method for desensitising an individual
CC to a selected polypeptide antigen. The method comprises administering a
CC composition that contains polypeptide antigens in an amount that
CC generates in the individual a state of hyporesponsiveness to the antigen
CC to allow desensitisation to one or more polypeptide antigens. The method
CC of the invention has immunomodulator activity, and may have a use in gene
CC therapy. The composition and method are useful in manufacturing a
CC medicament for desensitising an individual to a selected polypeptide
CC antigen or for generating in the individual a state of hyporesponsiveness
CC to the antigen to allow desensitisation to one or more polypeptide
CC antigens. The present sequence is used in the exemplification of the
CC invention.
XX SQ Sequence 213 AA;
Query Match 100.0%; Score 213; DB 7; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e-203; Mismatches 0; Indels 0; Gaps 0;
Matches 213; Conservative 0;

QY 1 MMKFLIIAFAVFAVSADPIHYDKITEINKAIDDAIAAIEQSETIDPMKVPDHADKFER 60
Db 1 MMKFLIIAFAVFAVSADPIHYDKITEINKAIDDAIAAIEQSETIDPMKVPDHADKFER 60
QY 61 HVGIVDFKGLAARNTEARGLKMKQKQGDANVKGEIVKAHLICVHDDIVSMEDLAY 120
Db 61 HVGIVDFKGLAARNTEARGLKMKQKQGDANVKGEIVKAHLICVHDDIVSMEDLAY 120
QY 121 KLGDLPHTTHVTSIDQFVVALSLEISDEGNITMTSFEVRQFANVNVNHIGLSILDPIFG 180
Db 121 KLGDLPHTTHVTSIDQFVVALSLEISDEGNITMTSFEVRQFANVNVNHIGLSILDPIFG 180
QY 181 VLSDLVLTAFQDITVRKEMTKVLAPAPKRELEKN 213
Db 181 VLSDLVLTAFQDITVRKEMTKVLAPAPKRELEKN 213
```

```
RESULT 4
AAR60575
ID AAR60575 standard; protein; 215 AA.
XX AC AAR60575;
XX AC AAR60575;
DT 25-MAR-2003 (revised)
DT 01-APR-1995 (first entry)
XX DE
XX DE House dust mite allergen DerpVII cDNA.
XX KW DerpVII allergen; antiallergic; allergy diagnosis.
XX OS Dermatophagoides pteronyssinus.
XX PN WO9420614-A1.
XX PD 15-SEP-1994.
XX PF 11-MAR-1994; 94WO-AU000117.
XX PR 12-MAR-1993; 93US-00031141.
XX PR 22-JUN-1993; 93US-00081540.
XX PA (CHIL-) INST CHILD HEALTH RES.
XX PI Thomas WR, Chua K;
XX WPI; 1994-303021/37.
XX N-PSDB; AAQ71400.
XX New nucleic acid encoding specific dust mite allergens - and related
PT vectors, transformed cells, peptides and antibodies, useful for
PT desensitisation and diagnosis.
XX PS Claim 7; Page 36-37; 67pp; English.
XX CC DerpVII antigen is useful as antiallergic reagent for treating
CC sensitivity to house dust mite allergens. The DNA can be used as a probe
CC to detect the sensitivity of an individual to the allergen. (Updated on 25
CC -MAR-2003 to correct PN field.)
XX SQ Sequence 215 AA;
Query Match 11.7%; Score 25; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.2e-16; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

QY 155 TSFEVROFANVNVNHIGLSILDPIF 179
Db 155 TSFEVROFANVNVNHIGLSILDPIF 179
RESULT 5
AAY25586
ID AAY25586 standard; protein; 215 AA.
XX AC AAY25586;
XX AC AAY25586;
DT 30-SEP-1999 (first entry)
XX DE
XX DE D. pteronyssinus allergen Der p 7 protein fragment.
XX KW Major histocompatibility complex; class II; desensitising; human;
XX KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;
XX KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
XX KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
XX KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
XX KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX OS Dermatophagoides pteronyssinus.
```

XX WO9934826-A1.
 XX PN
 XX 15-JUL-1999.
 XX PF
 XX 11-JAN-1999; 99WO-GB000080.
 XX PR 09-JAN-1998; 98GB-00000445.
 XX PR 21-SEP-1998; 98GB-00020474.
 XX PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
 XX PI Larche M, Kay AB;
 XX WPI; 1999-458255/38.
 XX DR Desensitising patients to polypeptide allergens.
 XX PS Example 6; Page 51; 117pp; English.
 XX CC This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitising patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, cockroach, larvae of tenbrito molitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents the house dust mite (Dermatophagoides pteronyssinus) allergen Der p 7
 XX SQ Sequence 215 AA;
 Query Match 11.7%; Score 25; DB 2; Length 215;
 Best Local Similarity 100.0%; Pred. No. 2.2e-16;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 TSFEVRQFANVNVHIGLSILDPIF 179
 DB 155 TSFEVRQFANVNVHIGLSILDPIF 179
 RESULT 6
 ADC34836
 ID ADC34836 standard; protein; 215 AA.
 XX AC ADC34836;
 XX DT 18-DEC-2003 (first entry)
 XX DE House dust mite allergen Der p 7.
 XX KW house dust mite; allergen; antigen; hyporesponsive; desensitisation;
 XX KW immunomodulator; gene therapy.
 XX OS Dermatophagoides pteronyssinus.
 XX PN WO2003047618-A2.
 XX PD 12-JUN-2003.
 XX PF 05-DEC-2002; 2002WO-GB005548.
 XX PR 05-DEC-2001; 2001US-0338385P.
 XX PA (CIRC-) CIRCASSIA LTD.

XX Larche M, Ledger PW;
 XX WPI; 2003-523267/49.
 XX PT Desensitizing an individual to a selected polypeptide antigen comprises administering a composition containing polypeptide antigens in an amount that generates a state of hyporesponsiveness to the antigen to allow desensitization.
 XX PS Disclosure; Page 21; 57pp; English.
 XX CC The invention relates to a novel method for desensitising an individual to a selected polypeptide antigen. The method comprises administering a composition that contains polypeptide antigens in an amount that generates in the individual a state of hyporesponsiveness to the antigen to allow desensitisation to one or more polypeptide antigens. The method of the invention has immunomodulator activity, and may have a use in gene therapy. The composition and method are useful in manufacturing a medicament for desensitising an individual to a selected polypeptide antigen or for generating in the individual a state of hyporesponsiveness to the antigen to allow desensitisation to one or more polypeptide antigens. The present sequence is used in the exemplification of the invention.
 XX SQ Sequence 215 AA;
 Query Match 11.7%; Score 25; DB 7; Length 215;
 Best Local Similarity 100.0%; Pred. No. 2.2e-16;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 TSFEVRQFANVNVHIGLSILDPIF 179
 DB 155 TSFEVRQFANVNVHIGLSILDPIF 179
 RESULT 7
 ABP27647
 ID ABP27647 standard; protein; 87 AA.
 XX AC ABP27647;
 XX DT 02-JUL-2002 (first entry)
 XX DE Streptococcus polypeptide SEQ ID NO 4470.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX OS Streptococcus agalactiae.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB004789.
 XX PR 27-OCT-2000; 2000GB-00026333.
 XX PR 24-NOV-2000; 2000GB-00028727.
 XX PR 07-MAR-2001; 2001GB-00005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;
 XX PI Tettelin H;
 XX DR WPI; 2002-352536/38.
 XX DR N-PSDB; ABN68278.
 XX PT New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for

PT detecting a compound that binds to the protein.
XX Claim 1; Page 3608; 4525pp; English.
PS
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (i), nucleic acids encoding (i), ABN66044-ABN71526 and
CC antibodies that bind (i) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (i) are used to detect Streptococcus in a
CC biological sample. (i) is used to determine whether a compound binds to
CC (i). A composition comprising (i) or a nucleic acid encoding (i), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (i) may be used to recombinantly produce (i) and may be
CC used in gene therapy. Antibodies to (i) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 87 AA;

Query Match 3.8%; Score 8; DB 5; Length 87;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDAAIAIE 41
Db 71 DDAAIAIE 78
|||||
|

RESULT 8
AAG31332
ID AAG31332 standard; protein; 321 AA.
XX
AC AAG31332;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37610.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PP 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.

```
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148585P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150894P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159231P.
PR 13-OCT-1999; 99US-0159234P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.

PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.8%; Score 8; DB 3; Length 321;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QDFVVALS 143
   |||||
Db 304 QDFVVALS 311

RESULT 9
AAG31331
ID AAG31331 standard; protein; 342 AA.
XX
AC AAG31331;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37609.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
```

PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142377P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144322P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145324P.
PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 06-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147835P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148584P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 3.8%; Score 8; DB 3; Length 342;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 QDFVVALS 143
DB 325 QDFVVALS 332
RESULT 10
AAG31330
ID AAG31330 standard; protein; 349 AA.
AC AAG31330;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37608.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; Gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
FD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140981P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

```
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0149368P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.

PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.8%; Score 8; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QDFVVALS 143
DB 332 QDFVVALS 339

RESULT 11
AAR41823
ID AAR41823 standard; protein; 412 AA.
XX AAR41823;
XX 24-OCT-2003 (revised)
DT 23-MAR-1994 (first entry)
XX Methylobacillus glycogenes 1006 homoserine dehydrogenase.
XX L-threonine biosynthesis; homoserine dehydrogenase; threonine synthase;
KW homoserine kinase; amino acid; ATCC 21371.
XX Methylobacillus glycogenes; strain 1006 (ATCC 21371).
XX JP05207886-A.
PD 20-AUG-1993.
PF 29-JAN-1992; 92JP-00014335.
XX 29-JAN-1992; 92JP-00014335.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX WPI: 1993-297465/38.
DR N-PSDB; AAQ48930.
XX Prepn. of L-threonine by fermentation in methanol soln. - using
PT Methylobacillus sp. contg. recombinant vector to produce homoserine-
PT dehydrogenase or kinase, or threonine synthase.
XX Claim 5; Page 14-15; 21pp; Japanese.
XX DNA coding for at least one of the enzymes homoserine dehydrogenase,
CC threonine synthase and homoserine kinase is incorporated into a
CC recombinant expression vector and used to transform Methylobacillus
CC bacteria. The transformants are cultured in medium containing methanol as
CC carbon source to synthesise L-threonine. See also AAQ48929 and AAQ48931-
CC Q48932. (Updated on 24-OCT-2003 to standardise OS field)
XX Sequence 412 AA;

Query Match 3.8%; Score 8; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDAAIAAE 41
DB 386 DDAAIAAE 393

RESULT 12
ABU23355
ID ABU23355 standard; protein; 484 AA.
XX ABU23355;
XX 19-JUN-2003 (first entry)
XX
```

DE Protein encoded by Prokaryotic essential gene #8882.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bordetella pertussis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACA27225.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 51279; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation; or the activity of a gene in an operon required for

XX the gene product or that has an activity against a biological pathway

XX required for proliferation, or that inhibits cellular proliferation; (8)

XX identifying a gene required for cellular proliferation or the biological

XX pathway in which a proliferation-required gene or its gene product lies

XX or a gene on which the test compound that inhibits proliferation of an

XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX compound's activity; (11) a culture comprising strains in which the gene

XX product is overexpressed or underexpressed; (12) determining the extent

XX to which each of the strains is present in a culture or collection of

XX strains; or (13) identifying the target of a compound that inhibits the

XX Query Match

XX Best Local Similarity 3.8%; Score 8; DB 6; Length 484;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 76 IEARGLKQ 83

XX DB 415 IEARGLKQ 422

RESULT 13

AAW68402

ID AAW68402 standard; protein; 580 AA.

XX AAW68402;

XX 28-OCT-1998 (first entry)

XX FeLV-A gag protein.

XX Multivalent vaccine; cat; pathogen; respiratory disease; FeLV; FPV; FCV;

XX digestive disease; feline leukemia virus; feline pan-leukemia virus;

XX feline calicivirus; feline immunodeficiency virus; FIV; rabies virus;

XX vector; envelope glycoprotein.

XX Feline leukemia virus.

XX FR2751223-A1.

XX 23-JAN-1998.

XX 19-JUL-1996; 96FR-00009337.

XX 19-JUL-1996; 96FR-00009337.

XX (INNR) RHONE MERIEUX SA.

XX Audonnet JCF, Bouchardon A, Baudu P, Riviere MEA;

XX WPI; 1998-112823/11.

XX N-PSDB; AAV49262.

XX Multi-valent polynucleotide vaccines against feline pathogens - consist

XX of at least 3 plasmids able to express protective antigens from specified

XX viruses.

XX Example 9; Fig 5; 42pp; French.

XX The invention relates to a multivalent vaccine for protecting cats

XX against several pathogens, especially pathogens associated with

XX respiratory and digestive diseases. The pathogens are especially selected

XX from feline leukemia virus (FeLV), feline pan-leukemia virus (FPV),

XX feline calicivirus (FCV), feline immunodeficiency virus (FIV), or rabies

XX virus. The vaccines are preferably composed of polynucleotide sequences

XX encoding 3 antigens, all as part of vectors. This sequence represents the

XX gag protein from the feline leukemia virus (FeLV) subtype A strain

XX Glasgow-1. The coding sequence was sub-cloned into the plasmid pVR1012 to

XX generate plasmid pPB181 for use in the vaccine

XX Sequence 580 AA;

XX Query Match 3.8%; Score 8; DB 2; Length 580;

XX Best Local Similarity 100.0%; Pred. No. 45;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203

DB 513 KEMTKVLA 520

RESULT 14

AAW50122

ID AAW50122 standard; protein; 611 AA.

XX AAW50122;

XX 21-DEC-2001 (first entry)

XX Feline leukemia virus p27-gp70 recombinant antigen pFeLVp27-gp70611.

XX FeLV; glycoprotein; gp70; p27; pFeLVp27-gp70611; antigen; immune status;

XX vaccination status; cat.
 XX Feline herpesvirus.
 OS Synthetic.
 XX WO200166568-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 07-MAR-2001; 2001WO-US007251.
 XX
 XX 09-MAR-2000; 2000US-00521738.
 XX
 PA (HESKA-) HESKA CORP.
 PA (COLS) UNIV COLORADO STATE RES FOUND.
 XX
 PI Jensen WA, Lappin MR, Rosen DK, Andrews JS;
 XX
 XX WPI; 2001-639000/73.
 DR N-PSDB; AAH27069.
 XX
 XX Determining immune status or vaccination status of an animal to e.g.
 PT calicivirus comprises using a recombinant viral antigen.
 XX
 PS Claim 4; Page 119-121; 132pp; English.
 XX
 CC The present sequence is that of a fusion protein, termed pFelVp27-
 CC gp70611, between the C-terminus of feline leukemia virus (FeLV) pR65-gag
 CC and FeLV glycoprotein 70. The fusion protein was produced in Escherichia
 CC coli cells transformed by a recombinant vector comprising nucleic acid
 CC nFelVp27-gp701833 (see AAH27069). pFelVp27-gp70611 is an example of a
 CC recombinant infectious agent antigen that can be used in the method of
 CC the invention to determine the immune status of an animal. The method
 CC involves contacting a biological specimen of an animal (cat, dog or
 CC horse) with a recombinant antigen, and detecting the presence or absence
 CC of a complex between the recombinant antigen and an antibody present in
 CC the sample. The method determines whether the animal is protected against
 CC disease or should be vaccinated. Recombinant antigens (see AAH27069),
 CC nucleic acids encoding them (see AAH27069), methods of producing them,
 CC and assay methods are provided
 XX
 SQ Sequence 611 AA;
 Query Match 3.8%; Score 8; DB 4; Length 611;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 196 KEMTKVLA 203
 DB 131 KEMTKVLA 138
 |||||
 RESULT 15
 ABB92052
 ID ABB92052 standard; protein; 783 AA.
 XX
 AC ABB92052;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 XX Herbicide; plant; agriculture; herbicide.
 XX
 DE Herbicide; plant; agriculture; herbicide.
 XX
 XX Arabidopsis thaliana.
 XX
 XX WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 XX 28-AUG-2001; 2001WO-EP009892.
 XX
 XX 28-AUG-2001; 2001WO-EP009892.
 XX

XX (FARB) BAYER AG.
 PA
 XX Tietjen K, Weidner M;
 PI
 XX WPI; 2002-269010/31.
 DR
 XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.
 XX
 PS Claim 5; SEQ ID NO 1263; 261pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides
 XX
 SQ Sequence 783 AA;
 Query Match 3.8%; Score 8; DB 5; Length 783;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 KAIDDAIA 38
 DB 257 KAIDDAIA 264
 |||||
 RESULT 16
 AAR05898
 ID AAR05898 standard; protein; 1784 AA.
 XX
 AC AAR05898;
 XX
 DT 24-OCT-2003 (revised)
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 29-NOV-1990 (first entry)
 XX
 XX Gene product of first ORF of virus capable of inducing viraemia in feline
 XX species encoded by clone 61E.
 DE
 DE Viraemia; leukaemia; FeLV-A; clone 61C; clone 61E; ds.
 XX
 XX Feline leukemia virus; strain A.
 OS
 XX EP377842-A.
 XX
 PD 18-JUL-1990.
 XX
 PF 12-DEC-1989; 89EP-00122964.
 XX
 PR 13-DEC-1988; 88US-00284139.
 XX
 XX (HARD) UNIV HARVARD.
 PA (COLS) COLORADO STATE UNIV RES..
 PA (HARD) UNIV HARVARD.
 XX
 XX Hoover EA, Mullins JI;
 PI
 XX WPI; 1990-218326/29.
 DR N-PSDB; AAQ05252.
 XX
 XX DNA encoding leukaemia virus-A sub: type - producing AIDS type disease in
 PT cats, used to test drugs and vaccines.
 XX

PS Claim 6; Fig 1.1-1.7; 23pp; English.

XX Variant of viral genome induces immunodeficiency in cats similar to AIDS.
 CC May be used in research, especially in testing drugs and vaccines against
 CC viraemia and and feline leukaemia viruses. (Updated on 25-MAR-2003 to
 CC correct PA field.) (Updated on 27-AUG-2003 to correct OS field.) (Updated
 CC on 24-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1784 AA;

Query Match 3.8%; Score 8; DB 2; Length 1784;

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203
 |||||
 Db 513 KEMTKVLA 520

RESULT 17

AA94427
 ID AAR94427 standard; protein; 1784 AA.

XX AAR94427;

AC 16-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 11-JUN-1996 (first entry)

XX FELV FcA provirus clone 61E encoded nucleocapsid protein.
 DE
 XX
 XX FelV; retrovirus; vaccine; AIDS; disease model; immunodeficiency;
 KW viraemia; leukaemia; therapy; nucleocapsid.
 XX
 XX Feline immunodeficiency virus; subtype A.

Key Location/Qualifiers

FT Misc-difference 581

FT /note= "codon 581 in encoding sequence is TAG"

XX EP699758-A1.

XX PD 06-MAR-1996.

XX PF 12-DEC-1989; 95EP-00100871.

XX PR 13-DEC-1988; 88US-00284139.

XX (HARD) UNIV HARVARD.

XX (COLS) UNIV COLORADO STATE RES FOUND.

XX Hoover EA, Mullins JI;

XX WPI; 1996-180826/19.

XX N-PSDB; AAT13265.

XX Inactivated FeLV-A sub-type isolates for use in disease models and
 PT vaccines - can be used to study prophylaxis and therapy of related immuno
 PT -deficiencies in other species, e.g. human.
 XX
 PS Disclosure; Fig 1; 22pp; English.

XX The nucleocapsid protein (AAR94427) and envelope protein (AAR94428)
 CC sequences of feline leukaemia virus provirus FcA clone 61E were deduced
 CC from open reading frames identified in the proviral DNA (AAT13265). Clone
 CC 61E is replication competent and capable of inducing persistent viraemia
 CC in cats. Host cells, e.g. feline cells, can be transformed with DNA
 CC derived from 61E and used to produce infectious virus useful in vaccines,
 CC in the generation of viraemia and in disease challenge systems. (Updated
 CC on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to
 CC standardise OS field)

XX Sequence 1784 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203
 |||||
 Db 513 KEMTKVLA 520

RESULT 18

AA51943
 ID AAB51943 standard; protein; 73 AA.

XX AAB51943;

XX AC 21-FEB-2001 (first entry)

XX DE Human secreted protein sequence encoded by gene 15 SEQ ID NO:75.

XX Human; secreted protein; immunosuppressive; antiarthritic; vulnary;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; hyperproliferative disorder; neoplasm;
 KW autoimmune disease; rheumatoid arthritis; cardiovascular disorder;
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis; viral;
 KW nervous system disorder; Alzheimer's disease; bacteria; infection;
 KW cell proliferation; skin aging; wound healing; chemotaxis; food additive.

XX Homo sapiens.

XX WO2000050334-A1.

XX 05-OCT-2000.

XX 22-MAR-2000; 2000WO-US007507.

XX 26-MAR-1999; 99US-0126594P.

XX 17-DEC-1999; 99US-0172408P.

XX (HUYA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-611701/58.

XX N-PSDB; AAC95476.

XX New nucleic acid molecules encoding 50 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 PS Claim 11; Page 374; 402pp; English.

XX Polynucleotide sequences AAB95462 - AAC95511 represent cDNA encoding 50
 CC human secreted proteins AAB51929 - AAB51978. Sequences AAB51979 -
 CC AAB52010 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnary. The secreted
 CC proteins, polypeptides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. Oligonucleotides AAC95453 - AAC95461 and peptide AAB51928 are
 CC used in the isolation and characterisation of the proteins and
 CC polynucleotides of the invention
 XX
 XX Sequence 73 AA;

Query Match 3.3%; Score 7; DB 3; Length 73;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 TGFVRQ 161

Db 43 TGFVRQ 49

RESULT 19
 ID ABR41325
 AC ABR41325;
 XX ABR41325;

DT 02-JUN-2003 (first entry)

DE Human DTHP intracellular signalling protein.

XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
 KW cancer; cell proliferative disorder; autoimmune disorder;
 KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
 KW neurological disorder; gastrointestinal disorder; transport disorder;
 KW connective tissue disorder; drug screening; proteome analysis;
 KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
 KW disease model; toxicological testing; transcript imaging;
 KW intracellular signalling.

XX Homo sapiens.

OS
 PN WO200297031-A2.

XX 05-DEC-2002.

XX 27-MAR-2002; 2002WO-US010056.

XX 28-MAR-2001; 2001US-0279619P.

PR 29-MAR-2001; 2001US-0280067P.

PR 29-MAR-2001; 2001US-0280068P.

PR 16-MAY-2001; 2001US-0291280P.

PR 17-MAY-2001; 2001US-0291829P.

PR 17-MAY-2001; 2001US-0291849P.

PR 19-JUN-2001; 2001US-0299428P.

PR 20-JUN-2001; 2001US-0299776P.

PR 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anehey SR;

PI Daugherty SC, Dam TC, Liu IF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RV, Urashka NE;

XX WPI; 2003-129518/12.

DR N-PSDB; ACC46265.

XX Novel human diagnostic and therapeutic polypeptide useful for identifying
 PT test compound which specifically binds to a polypeptide encoded by human
 PT diagnostic and therapeutic polynucleotide, and to induce antibodies.

XX Claim 27; SEQ ID NO 860; 591pp; English.

XX The invention relates to novel human diagnostic and therapeutic

CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
 CC proteins (DTHP; ABR41136-ABR41812). The invention also relates to
 CC polynucleotide sequences at least 90% identical to the dithp cDNA
 CC sequences of the invention; recombinant vectors, host cells and
 CC transgenic organisms comprising a dithp nucleic acid sequence; the
 CC recombinant production of DTHP proteins; antibodies specific for DTHP
 CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
 CC detecting dithp nucleotide and protein sequences; methods of screening
 CC for compounds which specifically bind a DTHP protein; and methods of
 CC assessing the toxicity of test compounds using a dithp hybridisation
 CC probe. Dithp nucleic acid sequences and DTHP proteins may be used in the
 CC diagnosis of a wide variety of conditions including cancer and other cell
 CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
 CC viral, fungal or parasitic infections; hormonal disorders; metabolic
 CC disorders; neurological disorders; gastrointestinal disorders; transport
 CC disorders; and connective tissue disorders. They may also be used to
 CC screen for modulators of protein activity or gene expression. DTHP
 CC proteins can additionally be used in analysis of the proteome of a tissue
 CC or cell type and to induce antibodies. The dithp nucleic acids are
 CC additionally useful in somatic or germline gene therapy of the disorders
 CC mentioned above, as a source of antisense sequences, as a source of
 CC probes and primers, in genotyping and identification of individuals, in
 CC the generation of transgenic animal models of human disease or knock in
 CC humanised animals, in toxicological testing, and in transcript imaging.
 CC The present sequence represents a DTHP protein which has intracellular
 CC signalling activity. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 85 AA;

Query Match 3.3%; Score 7; DB 6; Length 85;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 DFKGELA 72

Db 30 DFKGELA 36

RESULT 20

AAU37687

ID AAU37687 standard; protein; 87 AA.

XX AAU37687;

XX 14-FEB-2002 (first entry)

DT Streptococcus pneumoniae cellular proliferation protein #116.

DE Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.

XX Streptococcus pneumoniae.

OS WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX PS Claim 1; SEQ ID NO 2354; 56pp; English.

XX CC The invention relates to a protein comprising or having at least 50%

XX CC identity to any of the 2469 amino acid sequences, identified in the

XX CC specification (available on a computer readable format), or its fragment,

XX CC expressed from 2469 of 2489 identified DNA coding regions from the

XX CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as

XX CC ABS56454. Also included are an antibody which binds one of the proteins,

XX CC treating a patient by administering the protein, DNA or antibody (in a

XX CC composition), a kit comprising first and second primers, which are the

XX CC nucleic acid cited above or fragments between nucleotides 8-100 of a

XX CC sequence not defined in the specification, for amplifying a target

XX CC sequence contained within a Streptococcus nucleic acid sequence, where

XX CC the first primer is substantially complementary to the target sequence

XX CC and the second primer is substantially complementary to the complement of

XX CC the target sequence, and where the parts of the primers having

XX CC substantial complementarity define the termini of the target sequence to

XX CC be amplified, assay comprising contacting a test compound with the

XX CC protein, and determining whether the test compound binds to the protein

XX CC and a Streptococcus pneumoniae bacterium, where one or more genes

XX CC encoding the proteins has been rendered inactive. The proteins, nucleic

XX CC acid molecules, antibody and compositions are useful as medicaments for

XX CC treating or preventing a disease or infection due to streptococcus

XX CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis

XX CC media or ear infection. They are also useful in developing vaccines,

XX CC diagnostics and antibiotics. The methods are useful for identifying

XX CC immunodominant proteins. The present sequence is one of the 2469 proteins

XX CC expressed by the identified coding regions from the genomic sequence.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to

XX CC standardise OS field)

XX CC Sequence 87 AA;

QY 34 DDAIAAI 40

Db 71 DDAIAAI 77

RESULT 23

ABU46065

ID ABU46065 standard; protein; 87 AA.

XX AC ABU46065;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by prokaryotic essential gene #31592.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Streptococcus pneumoniae.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362693P.

XX PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA49935.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 73989; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation or the activity of a gene in an operon required for

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC required for proliferation in cells other than S. aureus, S. typhimurium,

XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX CC Sequence 87 AA;

QY 34 DDAIAAI 40

Db 71 DDAIAAI 77

RESULT 24

ABU44690

ID ABU44690 standard; protein; 87 AA.

XX AC ABU44690;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by prokaryotic essential gene #30217.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Streptococcus mutans.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX

PF XX 21-MAR-2002; 2002WO-US009107.
PR XX 21-MAR-2001; 2001US-00815242.
PR XX 06-SEP-2001; 2001US-00948993.
PR XX 25-OCT-2001; 2001US-0342923P.
PR XX 08-FEB-2002; 2002US-00072851.
PR XX 06-MAR-2002; 2002US-0362699P.
XX XX (ELIT-) ELITRA PHARM INC.
XX XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX XX WPI; 2003-029926/02.
XX XX N-PSDB; ACA48560.
XX XX New antisense nucleic acids, useful for identifying proteins or screening
XX XX PT for homologous nucleic acids, required for cellular proliferation to
XX XX PT isolate candidate molecules for rational drug discovery programs.
XX XX Claim 25; SEQ ID NO 72614; 1765pp; English.
XX XX The invention relates to an isolated nucleic acid comprising any one of
XX XX CC the 6213 antisense sequences given in the specification where expression
XX XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX XX CC polypeptide or its fragment whose expression is inhibited by the antisense
XX XX CC nucleic acid; (4) an antibody capable of specifically binding
XX XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX XX CC proliferation or the activity of a gene in an operon required for
XX XX CC proliferation; (7) identifying a compound that influences the activity of
XX XX CC the gene product or that has an activity against a biological pathway
XX XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX XX CC identifying a gene required for cellular proliferation or the biological
XX XX CC pathway in which a proliferation-required gene or its gene product lies
XX XX CC or a gene on which the test compound that inhibits proliferation of an
XX XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX XX CC compound's activity; (11) a culture comprising strains in which the gene
XX XX CC product is overexpressed or underexpressed; (12) determining the extent
XX XX CC to which each of the strains is present in a culture or collection of
XX XX CC strains; or (13) identifying the target of a compound that inhibits the
XX XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX XX CC identifying proteins or screening for homologous nucleic acids required
XX XX CC for cellular proliferation to isolate candidate molecules for rational
XX XX CC drug discovery programs, or for screening homologous nucleic acids
XX XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX XX CC patent did not form part of the printed specification, but was obtained
XX XX CC in electronic format directly from WIPO at
XX XX ftp.wipo.int/pub/published_pct_sequences
XX XX Sequence 87 AA;
XX XX Query Match 3.3%; Score 7; DB 6; Length 87;
XX XX Best Local Similarity 100.0%; Pred. No. 81;
XX XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 34 DDAIAAI 40
DB 71 DDAIAAI 77
RESULT 25
ABU46703
ID ABU46703 standard; protein; 87 AA.
XX AC ABU46703;
XX XX 19-JUN-2003 (first entry)
XX XX

DE XX Protein encoded by Prokaryotic essential gene #32230.
KW XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Streptococcus pyogenes.
XX XX WO200277183-A2.
XX XX 03-OCT-2002.
XX XX 21-MAR-2002; 2002WO-US009107.
XX XX 21-MAR-2001; 2001US-00815242.
XX XX 06-SEP-2001; 2001US-00948993.
XX XX 25-OCT-2001; 2001US-0342923P.
XX XX 08-FEB-2002; 2002US-00072851.
XX XX 06-MAR-2002; 2002US-0362699P.
XX XX (ELIT-) ELITRA PHARM INC.
XX XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX XX WPI; 2003-029926/02.
XX XX N-PSDB; ACA50573.
XX XX New antisense nucleic acids, useful for identifying proteins or screening
XX XX PT for homologous nucleic acids required for cellular proliferation to
XX XX PT isolate candidate molecules for rational drug discovery programs.
XX XX Claim 25; SEQ ID NO 74627; 1766pp; English.
XX XX The invention relates to an isolated nucleic acid comprising any one of
XX XX CC the 6213 antisense sequences given in the specification where expression
XX XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX XX CC polypeptide or its fragment whose expression is inhibited by the antisense
XX XX CC nucleic acid; (4) an antibody capable of specifically binding
XX XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX XX CC proliferation or the activity of a gene in an operon required for
XX XX CC proliferation; (7) identifying a compound that influences the activity of
XX XX CC the gene product or that has an activity against a biological pathway
XX XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX XX CC identifying a gene required for cellular proliferation or the biological
XX XX CC pathway in which a proliferation-required gene or its gene product lies
XX XX CC or a gene on which the test compound that inhibits proliferation of an
XX XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX XX CC compound's activity; (11) a culture comprising strains in which the gene
XX XX CC product is overexpressed or underexpressed; (12) determining the extent
XX XX CC to which each of the strains is present in a culture or collection of
XX XX CC strains; or (13) identifying the target of a compound that inhibits the
XX XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX XX CC identifying proteins or screening for homologous nucleic acids required
XX XX CC for cellular proliferation to isolate candidate molecules for rational
XX XX CC drug discovery programs, or for screening homologous nucleic acids
XX XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX XX CC patent did not form part of the printed specification, but was obtained
XX XX CC in electronic format directly from WIPO at
XX XX ftp.wipo.int/pub/published_pct_sequences
XX XX Sequence 87 AA;
XX XX Query Match 3.3%; Score 7; DB 6; Length 87;
XX XX Best Local Similarity 100.0%; Pred. No. 81;
XX XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 35 DAAIAIE 41
DB 72 DAAIAIE 78

RESULT 26

AA001133
ID AAG01133 standard; protein; 133 AA.

XX AC AAG01133;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein, SEQ ID NO: 5214.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX FN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-00200610.

XX PR 26-FEB-1999; 99US-0122487P.

XX PA (GIST) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX DR N-PSDB; AAC01139.

XX CC New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX CC obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX CC diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX PS Claim 13; SEQ ID NO 5214; 71pp + Sequence Listing; English.

XX CC The present sequence is a polypeptide encoded by one of a large number of
XX CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
XX CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
XX CC tissues. EST sequences usually correspond mainly to the 3' untranslated
XX CC region (UTR) of the mRNA because they are often obtained from oligo-dT
XX CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
XX CC sequences derived from the 5' ends of mRNAs and even in those cases where
XX CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
XX CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
XX CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
XX CC are also used in diagnostic, forensic, gene therapy and chromosome
XX CC mapping procedures. They are used to obtain upstream regulatory sequences
XX CC and to design expression and secretion vectors

XX SQ Sequence 133 AA;

Query Match 3.3%; Score 7; DB 3; Length 133;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19

DB 59 VAVSADP 65

|||||

RESULT 27

AAU30907
ID AAU30907 standard; protein; 139 AA.

XX AC AAU30907;

XX DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #1399.

XX

Human; vaccination; gene therapy; nutritional supplement;
stem cell proliferation; haematopoiesis; nerve tissue regeneration;
immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

XX WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US008656.

XX PR 18-APR-2000; 2000US-00552929.

XX PR 26-JAN-2001; 2001US-00770160.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-611725/70.

XX CC Nucleic acids encoding a range of human polypeptides, useful in genetic
XX CC vaccination, testing and therapy.

XX PS Claim 20; Page 368; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising the
XX CC nucleic acids encoding the polypeptides and cells genetically engineered
XX CC to express them are also useful for producing the proteins. The proteins
XX CC are useful in genetic vaccination, testing and therapy, and can be used
XX CC as nutritional supplements. They may be used to increase stem cell
XX CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
XX CC and/or nerve tissue growth or regeneration; immune suppression and/or
XX CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
XX CC AAU29510-AAU3304 represent the amino acid sequences of novel human
XX CC secreted proteins of the invention

XX SQ Sequence 139 AA;

Query Match 3.3%; Score 7; DB 4; Length 139;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19

DB 59 VAVSADP 65

|||||

RESULT 28

ADA33025
ID ADA33025 standard; protein; 148 AA.

XX AC ADA33025;

XX DT 20-NOV-2003 (first entry)

XX DE Acinetobacter baumannii protein #186.

XX KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX KW plant biocontrol agent.

XX OS Acinetobacter baumannii.

XX PN US6562958-B1.

XX XX 13-MAY-2003.

PD

```

XX PF 04-JUN-1999; 99US-00328352.
XX PF 09-JUN-1998; 98US-0088701P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton G, Bush D;
XX DR WPI; 2003-576092/54.
XX DR N-PSDB; ADA28899.
XX XX
XX PF New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX PT for diagnosing a bacterial disease, as components of antibacterial
XX PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX PT plants.
XX PS Example; SEQ ID NO 4312; 328pp; English.
XX CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX CC The A. baumannii nucleic acids and polypeptides are useful as reagents
XX CC for diagnosing a bacterial disease, as components of antibacterial
XX CC vaccines, as targets for antibacterial drugs, to detect the presence of
XX CC A. baumannii and other Acinetobacter species in a sample, in screening
XX CC compounds for the ability to interfere with the A. baumannii life cycle
XX CC or to inhibit A. baumannii infection, and as biocontrol agents for
XX CC plants. The present sequence represents the amino acid sequence of an A.
XX CC baumannii protein.
XX CC Sequence 148 AA;
XX SQ
XX Query Match 3.3%; Score 7; DB 6; Length 148;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 169 IGGLSIL 175
XX DB 21 IGGLSIL 27
XX RESULT 29
XX AAU30909
XX ID AAU30909 standard; protein; 152 AA.
XX AC AAU30909;
XX XX
XX DT 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #1400.
XX KW Human; vaccination; gene therapy; nutritional supplement;
XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX XX
XX FN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US008656.
XX XX
XX PR 18-APR-2000; 2000US-00552929.
XX PR 26-JAN-2001; 2001US-00770160.
XX PA (HYSE-) HYSEQ INC.
XX XX
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-611725/70.
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy.
XX PT
XX Claim 20; Page 369-370; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent or
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising the
XX CC nucleic acids encoding the polypeptides and cells genetically engineered
XX CC to express them are also useful for producing the proteins. The proteins
XX CC are useful in genetic vaccination, testing and therapy, and can be used
XX CC as nutritional supplements. They may be used to increase stem cell
XX CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
XX CC and/or nerve tissue growth or regeneration; immune suppression and/or
XX CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
XX CC AAU29510-AAU33304 represent the amino acid sequences of novel human
XX CC secreted proteins of the invention
XX XX
XX SQ Sequence 152 AA;
XX Query Match 3.3%; Score 7; DB 4; Length 152;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 13 VAVSADP 19
XX DB 27 VAVSADP 33
XX RESULT 30
XX AAG61764
XX ID AAG61764 standard; protein; 155 AA.
XX XX
XX AC AAG61764;
XX XX
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 80209.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX XX
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX XX
XX PF 25-FEB-2000; 2000EP-00301439.
XX XX
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126264P.
XX PR 29-MAR-1999; 99US-0126785P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129845P.
XX PR 19-APR-1999; 99US-0130077P.
XX PR 21-APR-1999; 99US-0130449P.
XX PR 23-APR-1999; 99US-0130510P.
XX PR 23-APR-1999; 99US-0130891P.
XX PR 28-APR-1999; 99US-0131449P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 04-MAY-1999; 99US-0132407P.
XX PR 04-MAY-1999; 99US-0132484P.
XX PR 05-MAY-1999; 99US-0132485P.

```

99US-0145089P
99US-0145124P
99US-0145145P
99US-0145218P
99US-0145242P
99US-0145267P
99US-0145313P
99US-0145318P
99US-0145319P
99US-0145361P
99US-0145368P
99US-0145388P
99US-0145398P
99US-0147038P
99US-0147204P
99US-0147302P
99US-0147129P
99US-0147260P
99US-0147303P
99US-0147416P
99US-0147493P
99US-0147935P
99US-0148171P
99US-0148313P
99US-0148344P
99US-0148565P
99US-0148684P
99US-0149368P
99US-0149175P
99US-0149426P
99US-0149723P
99US-0149729P
99US-0149923P
99US-0149929P
99US-0149930P
99US-0149930P
99US-0150566P
99US-0150884P
99US-0151062P
99US-0151068P
99US-0151080P
99US-0151303P
99US-0151438P
99US-0151930P
99US-0152363P
99US-0153070P
99US-0153070P
99US-0153758P
99US-0154018P
99US-0154039P
99US-0154779P
99US-0155139P
99US-0155486P
99US-0155659P
99US-0155659P
99US-0156458P
99US-0156596P
99US-0157517P
99US-0157753P
99US-0157865P
99US-0158029P
99US-0158232P
99US-0158369P
99US-0159293P
99US-0159293P
99US-0159294P
99US-0159329P
99US-0159329P
99US-0159330P
99US-0159331P
99US-0159637P
99US-0159638P
99US-0159584P
99US-0160741P
99US-0160767P
99US-0160768P
99US-0160770P

PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 22-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161408P.
 PR 25-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.
 PR 29-OCT-1999; 99US-0162143P.

Query Match 3.3%; Score 7; DB 3; Length 155;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAIFQ 191
 |||||
 DB 115 VLTAIFQ 121
 |||||

RESULT 31
 ABP26815
 ID ABP26815 standard; protein; 178 AA.
 AC ABP26815;
 XX
 XX
 DT 02-JUL-2002 (first entry)
 DE Streptococcus polypeptide SEQ ID NO 2806.
 DE Streptococcus GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX Streptococcus pyogenes.
 OS
 XX WO200234771-A2.
 FN
 XX
 PD 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB004789.
 XX
 XX 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX
 XX (CHIR-) CHIRON SPA.
 PA
 XX (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;
 PI Tettelin H;
 PR
 XX WPI; 2002-352536/38.
 DR N-PSDB; ABN67445.
 XX
 XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX
 PS Claim 1; Page 3431; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (i), nucleic acids encoding (i), ABN66044-ABN71526 and
 CC antibodies that bind (i) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (i) are used to detect Streptococcus in a
 CC biological sample. (i) is used to determine whether a compound binds to
 CC (i). A composition comprising (i) or a nucleic acid encoding (i), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (i) may be used to recombinantly produce (i) and may be
 CC used in gene therapy. Antibodies to (i) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 SQ Sequence 178 AA;
 Query Match 3.3%; Score 7; DB 5; Length 178;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 FGVLSDV 185
 |||||
 DB 123 FGVLSDV 129
 |||||

RESULT 32
 ADD71115
 ID ADD71115 standard; protein; 178 AA.
 XX
 AC ADD71115;
 XX
 DT 15-JAN-2004 (first entry)
 DE Human intracellular signalling molecule INTSIG-4 protein SEQ ID NO:4.
 DE human; intracellular signalling molecule; INTSIG; cytosolic;
 KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;
 KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
 KW antiinflammatory; thymimetic; gene therapy;
 KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;
 KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;
 KW immune disorder; inflammatory disorder; AIDS; allergy;
 KW developmental disorder; Hypothyroidism; Cushing's syndrome;
 KW reproductive disorder; vesicle-trafficking disorder; infection.
 XX
 XX Homo sapiens.
 OS
 XX WO2003039348-A2.
 FN
 XX 15-MAY-2003.
 PD
 XX 07-NOV-2002; 2002WO-US036151.
 XX
 XX 09-NOV-2001; 2001US-0344472P.
 PR 30-NOV-2001; 2001US-0334558P.
 PR 14-DEC-2001; 2001US-0340296P.
 PR 21-DEC-2001; 2001US-0343557P.
 PR 18-JAN-2002; 2002US-0350420P.
 PR 25-JAN-2002; 2002US-0351927P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Kable AB, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;
 PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;
 PI Marquis JP, Gietzen KU, Baughn MR, Borowsky ML, Yao MG, Chawla NK;
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;
 PI Lee SY, Tran UK, Elliott VS, Sprague WW, Tang YT, Zebbarjadian Y;
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;
 PI Zheng W;
 XX WPI; 2003-441441/41.
 DR N-PSDB; ADD71167.
 XX
 XX New human intracellular signalling molecules (INTSIG)), useful for
 PT diagnosing, treating and preventing diseases or conditions associated

PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or
infections.
XX Claim 1; SEQ ID NO 4; 363pp; English.
XX The present invention describes human intracellular signalling molecules
CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins
CC have cytosolic, antiapoptotic, antidiabetic, anticonvulsant,
CC neurotrophic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
CC antiinflammatory and thymic activities, and can be used in gene
CC therapy. The INTSIG polynucleotides and proteins can be used in the
CC diagnosis, treatment and prevention of diseases or conditions associated
CC with the decreased expression or overexpression of INTSIG, such as cell
CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),
CC neurological (e.g. epilepsy, Huntington's disease, stroke),
CC immune/inflammatory (e.g. AIDS, allergies), developmental (e.g.
CC hypothyroidism, Cushing's syndrome), reproductive and vesicle-trafficking
CC disorders, or infections. They are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of INTSIG. The INTSIG or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The present sequence represents human INTSIG-4 from the present
CC invention.
XX
XX
SQ Sequence 178 AA;
Query Match 3.3%; Score 7; DB 7; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 VAYSADP 19
Db 59 VAYSADP 65
RESULT 33
AAAG59755
ID AAG59755 standard; protein; 181 AA.
XX
AC AAG59755;
DT
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 77324.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 03-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.


```
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154019P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154019P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
Query Match 3.3%; Score 7; DB 3; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 VLTAIFQ 191
Db 83 VLTAIFQ 89
RESULT 34
AAG59754
ID AAG59754 standard; protein; 183 AA.
XX
AC AAG59754;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 77323.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
```

PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132456P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144844P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.

PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149436P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159232P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.


```
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147360P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149802P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Gaps 0;
Matches 7; Conservative 0; Indels 0;

QY 185 VLTAFQ 191
Db 87 VLTAFQ 93

RESULT 36
AAG55324
ID AAG55324 standard; protein; 185 AA.
XX
AC AAG55324;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 70920.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127482P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
```

PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140355P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140951P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 21-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148344P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 27-AUG-1999; 99US-0151303P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156598P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 VLTAIFQ 191
DB 88 VLTAIFQ 94
|||||

RESULT 37
ABP69706
ID ABP69706 standard; protein; 190 AA.
XX
AC ABP69706;
XX
DT 20-JAN-2003 (first entry)
XX
DE Human polypeptide SEQ ID NO 1753.
XX
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic.
XX
OS Homo sapiens.
XX
FN WO200270539-A2.
XX
PD 12-SEP-2002.
XX
PF 05-MAR-2002; 2002WO-US005095.
XX
PR 05-MAR-2001; 2001US-00799451.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR WPI; 2002-759812/82.
DR N-PSDB; ABZ11923.
XX
PT New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.
XX
PS Claim 9; SEQ ID NO 1753; 1012pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences (ABZ1119-
CC ABZ12066) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP6902-ABP69849) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burn, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 190 AA;

Query Match 3.3%; Score 7; DB 5; Length 190;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19
DB 59 VAVSADP 65
|||||

RESULT 38
ADA24313
ID ADA24313 standard; protein; 192 AA.
XX
AC ADA24313;
XX
DT 20-NOV-2003 (first entry)
XX
DE Bacillus clausii Sipv protein partial sequence.
XX
KW secretion factor; SF; SecA; SecD; SecE; SecF; SecG; SecY; Ffh; FtsY;
KW SipS; SipT; SipV; SipW; transcription gene; sporulation gene;
KW sporulation deficient; Gram positive microorganism; Bacillus genus;
KW large-scale industrial fermentation; fermentation product;
KW sporulation factor; industrial protein; therapeutic protein;
KW agricultural protein; food; animal feed; cosmetic; cleaning composition;
KW enzyme; hormone; growth factor.
XX
OS Bacillus clausii.
XX
FN WO2003066818-A2.
XX
PD 14-AUG-2003.
XX
PF 06-FEB-2003; 2003WO-US003534.
XX
PR 08-FEB-2002; 2002US-0355258P.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Ferrari E, Van Kimmenade A;
XX
DR WPI; 2003-646305/61.
DR N-PSDB; ADA24314.
XX
PT New nucleic acid from Bacillus clausii, useful for improving recombinant
PT protein production, encodes secretion, transcription or sporulation
PT factors, also encoded proteins.
XX
PS Claim 11; Fig 29; 67pp; English.
XX
CC This invention relates to novel DNA sequences which encode one of the
CC Bacillus clausii secretion factors (SF) SecA, SecD, SecE, SecF, SecG,
CC SecY, Ffh, FtsY, SipS, SipT, SipV and SipW. In addition, the invention
CC relates to novel transcription and sporulation genes in Bacillus clausii.
CC It is desirable that a production strain of Bacillus clausii should be
CC deficient in sporulation, hence a greater understanding of this process
CC is proposed by the identification of the genes of the invention. Gram
CC positive microorganisms, such as members of the genus Bacillus, have been
CC used for large-scale industrial fermentation due, in part, to their
CC ability to secrete their fermentation products into their culture media.
CC Increasing the ability of these microorganisms to express and secrete
CC polypeptides is therefore of great interest in many fields. The secretion
CC of polypeptides by members of the genus Bacillus may be further
CC understood and increased by making use of the proteins and DNA sequences
CC of the invention. Sequences from Bacillus clausii that encode
CC transcription or sporulation factors are used in recombinant expression
CC and secretion of proteins, for example those for industrial, therapeutic
CC or agricultural use, also in food, animal feed, cosmetics and cleaning
CC compositions, typically enzymes, hormones and growth factors.
CC Transformation with sequences that encode secretion, transcription or
CC sporulation factors may increase both production and secretion of
CC proteins in Bacillus. The present sequence is the partial amino acid
CC sequence of the Bacillus clausii Sipv protein of the invention.

XX SQ Sequence 192 AA;
 Query Match 3.3%; Score 7; DB 7; Length 192;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 MTSFEVR 160
 |||||
 DB 44 MTSFEVR 50

RESULT 39
 ABB06132
 ID ABB06132 standard; protein; 193 AA.
 XX AC ABB06132;
 DT 10-MAY-2002 (first entry)
 XX DE Human NS protein sequence SEQ ID NO:224.
 XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;
 KW anti-rheumatic; antiarthritic; antipsoriatic; ophthalmological; anti-HIV;
 KW vasotropic; antiarteriosclerotic; antiinflammatory; dermatological;
 KW anorectic; muscular; antiinfertility; cardiovascular; anticoagulant;
 KW antifibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiant;
 KW anticonvulsant; antidiabetic; tranquiliser; antidepressant; neurotropic;
 KW gastrointestinal; virucide; antiulcer; cancer; osteoporosis; dystonia;
 KW contraceptive; vaccine; gene therapy; multiple sclerosis; psoriasis;
 KW endometriosis; degenerative disease; atherosclerosis; glaucoma;
 KW rheumatoid arthritis; cataract; restenosis; muscular dystrophy; AIDS;
 KW inflammation; skin disorder; obesity; conglutination disease; hypertension;
 KW ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;
 KW diabetes; anxiety; depression; schizophrenia; viral disease; stroke;
 KW gastric ulcer; Alzheimer's disease.
 XX OS Homo sapiens.
 PN WO200206315-A2.
 XX PD 24-JAN-2002.
 XX PF 17-JUL-2001; 2001WO-IL000653.
 XX PR 18-JUL-2000; 2000IL-00137345.
 XX PR 15-DEC-2000; 2000IL-00140354.
 XX PA (COMP-) COMPUGEN LTD.
 XX PI Mintz L, Freilich S, Bernstein J;
 XX WPI: 2002-155037/20.
 XX DR N-PSDB; ABI39786.
 XX One hundred and twenty eight novel nucleic acid sequences, useful for
 treating and diagnosing e.g. cancer, asthma and Alzheimer's.
 PS Claim 6; Page 260-261; 290pp; English.
 CC ABL39691 to ABL39918 represent novel human nucleic acid sequences
 encoding the proteins given in ABB06037 to ABB06164. The novel sequences
 (NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,
 CC anti-rheumatic, antiarthritic, antipsoriatic, ophthalmological, virucide,
 CC vasotropic, antiarteriosclerotic, antiinflammatory, dermatological,
 CC anorectic, muscular, anti-HIV, antiinfertility, cardiovascular,
 CC anticoagulant, antifibrinolytic, hypotension, antiasthmatic, cardiant,
 CC immunomodulator, anticonvulsant, antidiabetic, tranquiliser, antiulcer,
 CC antidepressant, gastrointestinal, neurotropic, cerebroprotective,
 CC neurotropic and contraceptive activities. The NS can be used in vaccines,
 CC gene therapy and antisense therapy. Nucleic acids, expression vectors and
 CC antibodies from the present invention can be used for treating and

CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative
 CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
 CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
 CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
 CC disease, coagulation disease, ischaemia, hypertension, asthma, immune
 CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
 CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
 CC Alzheimer's disease and as a contraceptive
 XX SQ Sequence 193 AA;
 Query Match 3.3%; Score 7; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
 |||||
 DB 59 VAVSADP 65

RESULT 40
 AAR54114
 ID AAR54114 standard; protein; 204 AA.
 XX AC AAR54114;
 DT 25-MAR-2003 (revised)
 DT 20-JAN-1995 (first entry)
 XX DE GDI D4 protein.
 XX D4; GDP-dissociation inhibitor; GDI; hematopoietic cell; diagnostic;
 KW leukemia; tumor bone marrow transplant.
 XX OS Bos taurus.
 XX WO9413802-A1.
 XX PD 23-JUN-1994.
 XX PF 10-DEC-1993; 93WO-US012074.
 XX PR 10-DEC-1992; 92US-00990337.
 XX (CYTO-) CYTOMED INC.
 XX (BETH-) BETH ISRAEL HOSPITAL ASSOC.
 XX Lim B, Lelias J, Adra CN, Ko JL;
 XX WPI: 1994-217878/26.
 XX New D4 protein preferentially expressed in hematopoietic cells - for use
 as a diagnostic, in research and for treatment of leukaemia tumours, and
 in bone marrow transplants.
 PS Disclosure; Page 56-57; 76pp; English.
 XX Differential screening of a subtractive hematopoietic cDNA library
 yielded D4 cDNA, which encoded a protein that showed GDP-dissociation
 CC inhibitor activity. The cDNA sequence was determined (AAQ68715) and the
 CC amino acid sequence corresponding to the longest ORF was deduced
 CC (AAR54112). Equivalent sequences for mouse D4 are given in AAQ68716 and
 CC (AAR54113). The cattle D4 protein sequence was also provided (AAR54114).
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 204 AA;
 Query Match 3.3%; Score 7; DB 2; Length 204;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
 |||||

Db	59	VAVSADP 65	PR	18-JUN-1999;	99US-0139456P.
			PR	18-JUN-1999;	99US-0139457P.
			PR	18-JUN-1999;	99US-0139458P.
			PR	18-JUN-1999;	99US-0139459P.
			PR	18-JUN-1999;	99US-0139460P.
			PR	18-JUN-1999;	99US-0139461P.
			PR	18-JUN-1999;	99US-0139462P.
			PR	18-JUN-1999;	99US-0139463P.
			PR	18-JUN-1999;	99US-0139750P.
			PR	18-JUN-1999;	99US-0139763P.
			PR	21-JUN-1999;	99US-0139817P.
			PR	22-JUN-1999;	99US-0139899P.
			PR	23-JUN-1999;	99US-0140353P.
			PR	23-JUN-1999;	99US-0140354P.
			PR	24-JUN-1999;	99US-0140695P.
			PR	28-JUN-1999;	99US-0140823P.
			PR	29-JUN-1999;	99US-0140911P.
			PR	30-JUN-1999;	99US-0141287P.
			PR	01-JUL-1999;	99US-0141842P.
			PR	01-JUL-1999;	99US-0142154P.
			PR	02-JUL-1999;	99US-0142055P.
			PR	06-JUL-1999;	99US-0142390P.
			PR	08-JUL-1999;	99US-0142803P.
			PR	09-JUL-1999;	99US-0142920P.
			PR	12-JUL-1999;	99US-0142977P.
			PR	13-JUL-1999;	99US-0143542P.
			PR	14-JUL-1999;	99US-0143624P.
			PR	15-JUL-1999;	99US-0144005P.
			PR	16-JUL-1999;	99US-0144085P.
			PR	16-JUL-1999;	99US-0144086P.
			PR	19-JUL-1999;	99US-0144331P.
			PR	19-JUL-1999;	99US-0144332P.
			PR	19-JUL-1999;	99US-0144333P.
			PR	19-JUL-1999;	99US-0144334P.
			PR	19-JUL-1999;	99US-0144335P.
			PR	20-JUL-1999;	99US-0144352P.
			PR	20-JUL-1999;	99US-0144632P.
			PR	20-JUL-1999;	99US-0144884P.
			PR	21-JUL-1999;	99US-0144814P.
			PR	21-JUL-1999;	99US-0145086P.
			PR	21-JUL-1999;	99US-0145088P.
			PR	22-JUL-1999;	99US-0145085P.
			PR	22-JUL-1999;	99US-0145087P.
			PR	22-JUL-1999;	99US-0145089P.
			PR	23-JUL-1999;	99US-0145192P.
			PR	23-JUL-1999;	99US-0145145P.
			PR	23-JUL-1999;	99US-0145218P.
			PR	26-JUL-1999;	99US-0145224P.
			PR	26-JUL-1999;	99US-0145276P.
			PR	27-JUL-1999;	99US-0145913P.
			PR	27-JUL-1999;	99US-0145918P.
			PR	27-JUL-1999;	99US-0145919P.
			PR	28-JUL-1999;	99US-0145951P.
			PR	02-AUG-1999;	99US-0146386P.
			PR	02-AUG-1999;	99US-0146388P.
			PR	02-AUG-1999;	99US-0146389P.
			PR	03-AUG-1999;	99US-0147038P.
			PR	04-AUG-1999;	99US-0147204P.
			PR	05-AUG-1999;	99US-0147302P.
			PR	05-AUG-1999;	99US-0147192P.
			PR	05-AUG-1999;	99US-0147260P.
			PR	06-AUG-1999;	99US-0147303P.
			PR	06-AUG-1999;	99US-0147416P.
			PR	09-AUG-1999;	99US-0147493P.
			PR	09-AUG-1999;	99US-0147935P.
			PR	10-AUG-1999;	99US-0148171P.
			PR	11-AUG-1999;	99US-0148319P.
			PR	12-AUG-1999;	99US-0148341P.
			PR	13-AUG-1999;	99US-0148565P.
			PR	13-AUG-1999;	99US-0148684P.
			PR	16-AUG-1999;	99US-0149368P.
			PR	17-AUG-1999;	99US-0149175P.

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 77322.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0125788P.

25-MAR-1999; 99US-0126264P.

29-MAR-1999; 99US-0126785P.

01-APR-1999; 99US-0127462P.

06-APR-1999; 99US-0128234P.

08-APR-1999; 99US-0128714P.

16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.

21-APR-1999; 99US-0130449P.

23-APR-1999; 99US-0130510P.

23-APR-1999; 99US-0130891P.

28-APR-1999; 99US-0131449P.

30-APR-1999; 99US-0132048P.

30-APR-1999; 99US-0132407P.

04-MAY-1999; 99US-0132484P.

05-MAY-1999; 99US-0132485P.

06-MAY-1999; 99US-0132486P.

08-MAY-1999; 99US-0132487P.

07-MAY-1999; 99US-0132863P.

11-MAY-1999; 99US-0134256P.

14-MAY-1999; 99US-0134218P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134221P.

14-MAY-1999; 99US-0134370P.

18-MAY-1999; 99US-0134768P.

18-MAY-1999; 99US-0134941P.

20-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135353P.

24-MAY-1999; 99US-0135629P.

25-MAY-1999; 99US-0136021P.

27-MAY-1999; 99US-0136392P.

28-MAY-1999; 99US-0136782P.

01-JUN-1999; 99US-0137222P.

03-JUN-1999; 99US-0137528P.

04-JUN-1999; 99US-0137502P.

07-JUN-1999; 99US-0137724P.

08-JUN-1999; 99US-0138094P.

10-JUN-1999; 99US-0138540P.

10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.

16-JUN-1999; 99US-0139452P.

16-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139492P.

18-JUN-1999; 99US-0139454P.

18-JUN-1999; 99US-0139455P.

us-10-024-955-7.oligo.rag

Mon Aug 9 08:14:01 2004

PR	18-AUG-1999;	99US-0149426P.	AC	AAG55323;
PR	20-AUG-1999;	99US-0149722P.	XX	
PR	20-AUG-1999;	99US-0149723P.	DT	18-OCT-2000 (first entry)
PR	20-AUG-1999;	99US-0149929P.	XX	
PR	23-AUG-1999;	99US-0149902P.	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 70919.
PR	23-AUG-1999;	99US-0149930P.	XX	
PR	25-AUG-1999;	99US-0150566P.	KW	Protein identification; signal transduction pathway; metabolic pathway;
PR	26-AUG-1999;	99US-0150884P.	KW	hybridisation assay; Genetic mapping; gene expression control; promoter;
PR	27-AUG-1999;	99US-0151065P.	XX	termination sequence.
PR	27-AUG-1999;	99US-0151066P.	OS	Arabidopsis thaliana.
PR	27-AUG-1999;	99US-0151080P.	XX	
PR	30-AUG-1999;	99US-0151303P.	PN	EP1033405-A2.
PR	31-AUG-1999;	99US-0151438P.	XX	
PR	01-SEP-1999;	99US-0151930P.	XX	
PR	07-SEP-1999;	99US-0152363P.	PD	06-SEP-2000.
PR	10-SEP-1999;	99US-0153070P.	XX	
PR	13-SEP-1999;	99US-0153758P.	PF	25-FEB-2000; 2000EP-00301439.
PR	15-SEP-1999;	99US-0154018P.	XX	
PR	16-SEP-1999;	99US-0154039P.	PR	25-FEB-1999; 99US-0121825P.
PR	20-SEP-1999;	99US-0154779P.	PR	05-MAR-1999; 99US-0123180P.
PR	22-SEP-1999;	99US-0155139P.	PR	09-MAR-1999; 99US-0123548P.
PR	23-SEP-1999;	99US-0155486P.	PR	23-MAR-1999; 99US-0125788P.
PR	24-SEP-1999;	99US-0155653P.	PR	25-MAR-1999; 99US-0126264P.
PR	28-SEP-1999;	99US-0156458P.	PR	29-MAR-1999; 99US-0126785P.
PR	29-SEP-1999;	99US-0156596P.	PR	01-APR-1999; 99US-0127462P.
PR	04-OCT-1999;	99US-0157117P.	PR	06-APR-1999; 99US-0128234P.
PR	05-OCT-1999;	99US-0157533P.	PR	08-APR-1999; 99US-0128714P.
PR	06-OCT-1999;	99US-0157865P.	PR	16-APR-1999; 99US-0129845P.
PR	07-OCT-1999;	99US-0158029P.	PR	19-APR-1999; 99US-0130077P.
PR	08-OCT-1999;	99US-0158234P.	PR	21-APR-1999; 99US-0130449P.
PR	12-OCT-1999;	99US-0158369P.	PR	23-APR-1999; 99US-0130510P.
PR	13-OCT-1999;	99US-0159293P.	PR	28-APR-1999; 99US-0130891P.
PR	13-OCT-1999;	99US-0159294P.	PR	30-APR-1999; 99US-0132048P.
PR	14-OCT-1999;	99US-0159329P.	PR	30-APR-1999; 99US-0132407P.
PR	14-OCT-1999;	99US-0159330P.	PR	04-MAY-1999; 99US-0132484P.
PR	14-OCT-1999;	99US-0159333P.	PR	05-MAY-1999; 99US-0132485P.
PR	14-OCT-1999;	99US-0159637P.	PR	06-MAY-1999; 99US-0132487P.
PR	14-OCT-1999;	99US-0159638P.	PR	07-MAY-1999; 99US-0132863P.
PR	18-OCT-1999;	99US-0159584P.	PR	11-MAY-1999; 99US-0134256P.
PR	21-OCT-1999;	99US-0160741P.	PR	14-MAY-1999; 99US-0134218P.
PR	21-OCT-1999;	99US-0160767P.	PR	14-MAY-1999; 99US-0134219P.
PR	21-OCT-1999;	99US-0160768P.	PR	14-MAY-1999; 99US-0134221P.
PR	21-OCT-1999;	99US-0160770P.	PR	14-MAY-1999; 99US-0134370P.
PR	21-OCT-1999;	99US-0160814P.	PR	18-MAY-1999; 99US-0134768P.
PR	21-OCT-1999;	99US-0160815P.	PR	19-MAY-1999; 99US-0134941P.
PR	22-OCT-1999;	99US-0160980P.	PR	20-MAY-1999; 99US-0135124P.
PR	22-OCT-1999;	99US-0160981P.	PR	21-MAY-1999; 99US-0135353P.
PR	22-OCT-1999;	99US-0160989P.	PR	24-MAY-1999; 99US-0135629P.
PR	25-OCT-1999;	99US-0161404P.	PR	25-MAY-1999; 99US-0136021P.
PR	25-OCT-1999;	99US-0161405P.	PR	27-MAY-1999; 99US-0136392P.
PR	25-OCT-1999;	99US-0161406P.	PR	28-MAY-1999; 99US-0136782P.
PR	26-OCT-1999;	99US-0161359P.	PR	01-JUN-1999; 99US-0137222P.
PR	26-OCT-1999;	99US-0161360P.	PR	03-JUN-1999; 99US-0137528P.
PR	26-OCT-1999;	99US-0161361P.	PR	04-JUN-1999; 99US-0137502P.
PR	28-OCT-1999;	99US-0161920P.	PR	07-JUN-1999; 99US-0137724P.
PR	28-OCT-1999;	99US-0161992P.	PR	08-JUN-1999; 99US-0138094P.
PR	28-OCT-1999;	99US-0161993P.	PR	10-JUN-1999; 99US-0138540P.
PR	29-OCT-1999;	99US-0162142P.	PR	10-JUN-1999; 99US-0138847P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-JUN-1999; 99US-0139119P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUN-1999; 99US-0139452P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUN-1999; 99US-0139453P.
PR	29-OCT-1999;	99US-0162142P.	PR	17-JUN-1999; 99US-0139452P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139454P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139455P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139456P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139457P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139458P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139459P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139460P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139461P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999; 99US-0139462P.

Query Match 3.3%; Score 7; DB 3; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAFQ 191
Db 107 VLTAFQ 113

PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142350P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144056P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.

PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154773P.
PR 22-SEP-1999; 99US-0155133P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160983P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 185 VLTAIFQ 191
|||
Db 117 VLTAIFQ 123

RESULT 43
AAG55906

ID AAG55906 standard; protein; 215 AA.

XX

AC AAG55906;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 71760.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 08-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 23-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161320P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 28-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 215;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAIFQ 191
 Db 119 VLTAIFQ 125
 |||||

RESULT 44
 ADB64094
 ID ADB64094 standard; protein; 267 AA.

XX ADB64094;

XX 04-DEC-2003 (first entry)

XX Human protein encoded by clone BRWHE20183170.

XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
 cell regeneration; membrane protein; signal transduction-related protein;
 transcription-related protein; osteoporosis; neurological disease;
 cancer; tumour.

XX Homo sapiens.

XX EPI1308459-A2.

PN

XX 07-MAY-2003.
 XX 28-MAR-2002; 2002EP-00007401.
 XX 05-NOV-2001; 2001JP-00379298.
 XX 25-JAN-2002; 2002US-00350978.
 XX (HELI-) HELIX RES INST.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 WPI; 2003-450961/43.
 XX N-PSDB; ADB62124.
 XX New polynucleotides and polypeptides, useful for developing a diagnostic
 marker or medicines for regulation of their expression and activity, or
 as targets of gene therapy.

Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected
 from 1970 fully defined nucleotide sequences which encode novel
 polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 or its partial peptide, an antibody binding to the polypeptide or peptide
 of the polynucleotide, immunologically assaying the polypeptide or
 peptide of the polynucleotide by contacting the polypeptide or peptide
 with the antibody of the encoded protein, and observing the binding
 between the two, a transformant carrying the polynucleotide in an
 expressible manner and an antisense polynucleotide. The oligonucleotide
 is useful as a primer for synthesizing the polynucleotide, or as a probe
 for detecting the polynucleotide. The polynucleotides and encoded
 proteins are useful as pharmaceutical agents and many disease-related
 genes may be included in them, for developing a diagnostic marker or
 medicines for regulation of their expression and activity, or as targets
 of gene therapy. The genes are involved in tissue and/or cell
 regeneration. Membrane proteins, signal transduction-related proteins,
 transcription-related proteins, disease-related proteins and genes
 encoding them can be used as indicators for diseases (e.g. osteoporosis,
 neurological diseases, cancer, tumours). The cDNA may be used to regulate
 the activity or expression of the encoded protein to treat diseases. The
 sequence presented is a protein of the invention. Note: Some of the
 sequence data for this patent is not represented in the printed
 specification, but is based on sequence information supplied by the
 European Patent Office.

SQ Sequence 267 AA;

Query Match 3.3%; Score 7; DB 7; Length 267;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATAAI 40
 Db 45 DDATAAI 51
 |||||

RESULT 45

ADB85214

ID ADB85214 standard; protein; 295 AA.

XX ADB85214;

XX 04-DEC-2003 (first entry)

XX Rat chimaerin SEQ ID NO:95.

XX rat; streptozocin; kinase; phosphatase; ion channel protein; receptor;

KW transporter; G-protein coupled receptor; GPCR; DNA-binding proteins;

KW protease; enzyme; analgesic; gene therapy; pain; diabetes.

XX OS Rattus norvegicus.
 XX PN EP1284297-A2.
 XX PD 19-FEB-2003.
 XX PF 26-JUL-2002; 2002EP-00255228.
 XX PR 27-JUL-2001; 2001GB-00018354.
 XX PR 07-FEB-2002; 2002GB-00002880.
 XX PA (WARN) WARNER LAMBERT CO.
 XX PI Brooksbank RA, Dixon AK, Lee K, Pimnock RD;
 XX DR N-PSDB; ADB85215.
 XX DR WPI; 2003-364994/35.
 XX DR N-PSDB; ADB85215.
 XX PT Use of gene sequence that is down-regulated in response to streptozocin-
 PT induced diabetes, vector, host cell, animal, polypeptide and antibody, in
 PT screening of compounds for treating or diagnosing pain.
 XX PS Disclosure; Page 162-163; 256pp; English.
 XX CC The invention relates to a novel isolated gene sequence that is down-
 CC regulated in the spinal cord in response to streptozocin-induced
 CC diabetes, or comprising, hybridising or having at least 80% sequence
 CC identity to a sequence whose expression products are kinases,
 CC phosphatases, ion channel proteins, receptors, transporters, G-protein
 CC coupled receptor proteins, DNA-binding proteins, proteases or enzymes,
 CC given in the specification. A gene of the invention has analgesic
 CC activity, and may have a use in gene therapy. The gene sequences, vector,
 CC host cell, animal, polypeptide and antibody are useful for screening of
 CC compounds for diagnosing or treating pain. The kits are useful for
 CC simultaneous, separate or sequential detecting and/or quantifying down-
 CC regulation of a gene sequence in the spinal cord of a mammal in response
 CC to streptozocin-induced diabetes. The compound or pharmaceutical
 CC composition is useful as a medicament for treating or diagnosing pain.
 CC The present sequence represents a protein encoded by a gene of the
 CC invention.
 XX SQ Sequence 295 AA;
 Query Match 3.3%; Score 7; DB 7; Length 295;
 Best Local Similarity 100.0%; Pred.No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 IEARGLK 82
 DB 128 IEARGLK 134
 RESULT 46
 AAB18216
 ID AAB18216 standard; protein; 300 AA.
 XX AC AAB18216;
 XX DT 07-NOV-2000 (first entry)
 XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:73.
 XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX OS Plasmodium falciparum.
 XX PN WO200025728-A2.
 XX PD 11-MAY-2000.
 XX PF 05-NOV-1999; 99WO-US026796.

XX OS 05-NOV-1998; 98US-0107131P.
 XX PR (HOFF/) HOFFMAN S.
 XX PA (CARU/) CARUCCI D.
 XX PA (GARD/) GARDNER M.
 XX PA (VENT/) VENTER J C.
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX DR WPI; 2000-365347/31.
 XX DR Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.
 XX PS Disclosure; Page 171-172; 577pp; English.
 XX CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX SQ Sequence 300 AA;
 Query Match 3.3%; Score 7; DB 3; Length 300;
 Best Local Similarity 100.0%; Pred.No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 207 KRELEKN 213
 DB 214 KRELEKN 220
 RESULT 47
 ADB09304
 ID ADB09304 standard; protein; 343 AA.
 XX AC ADB09304;
 XX DT 20-NOV-2003 (first entry)
 XX DE Alloiococcus otitis antigenic protein SEQ ID NO:3244.
 XX KW Alloiococcus otitidis; antigenic protein; immunogenic; immunisation;
 XX KW gene therapy; Gram-positive bacterium; infection.
 XX OS Alloiococcus otitis.
 XX PN WO2003048304-A2.
 XX PD 12-JUN-2003.
 XX PF 25-NOV-2002; 2002WO-US036123.
 XX PR 29-NOV-2001; 2001US-0333777P.

PR 18-NOV-2002; 2002US-0426742P.
PA (AMHP) WYETH HOLDINGS CORP.
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX WPI; 2003-505284/47.
XX N-PSDB; ADB09303.
XX New Alloicoccus otitidis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX Claim 33; SEQ ID NO 3244; 1019pp; English.
XX The present invention describes an isolated polynucleotide (I) of
XX Alloicoccus otitidis genomic DNA, which encodes an antigenic protein.
XX Alloicoccus otitidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
XX expression vector comprising the novel isolated polynucleotide (I), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Alloicoccus otitidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Alloicoccus
XX otitidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polypeptide from the culture. (I) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Alloicoccus
XX otitidis. The present sequence represents an Alloicoccus otitidis
XX antigen protein from the present invention.
XX Sequence 343 AA;
SQ Query Match 3.3%; Score 7; DB 6; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 FVVALSL 144
Db 209 FVVALSL 215
|||||
RESULT 48
AAR33439
ID AAR33439 standard; protein; 354 AA.
XX AAR33439;
XX 17-DEC-2001 (revised)
DT 06-JUL-1993 (first entry)
XX Ornithine cyclodeaminase C58 from Ti plasmid pTIC58.
XX mu-crystallins; drug targeting; nervous acting drugs; CNS; neural;
KW neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
KW neuromuscular junctions; memory agents; Alzheimers disease;
KW CNS depressants; CNS stimulators; tranquilizers; muscle relaxants;
KW antispasmodics; analgesics; anesthetics; anticonvulsants;
KW antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
KW hypnotics.
XX Agrobacterium tumefaciens.
OS

XX USN7844304-N.
PN 01-JAN-1993.
XX 28-FEB-1992; 92US-00844304.
PF 28-FEB-1992; 92US-00844304.
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
PA Wistow G, Kim R;
XX WPI; 1993-093573/11.
DR New mu-crystalline proteins - having ornithine cyclo-deaminase activity,
XX used in diagnosis and treatment of disorders in ornithine metabolism.
XX Disclosure; Page 34; 60pp; English.
XX This sequence represents ornithine cyclodeaminase (OCD) from
XX Agrobacterium Ti plasmid pTIC58. It shows approximately 30% homology with
XX the kangaroo eye lens protein mu-crystallin. (Note: Revised entry
XX submitted to correct the patent number format of US Government-owned NTIS
XX applications to prevent clashes with ongoing US granted patent numbers.
XX For further information please visit the Derwent web site at
XX www.derwent.com/dwpi/updates/ntis_us.html.)
XX Sequence 354 AA;
SQ Query Match 3.3%; Score 7; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 179 FGVLSDV 185
Db 96 FGVLSDV 102
|||||
RESULT 49
AAR33440
ID AAR33440 standard; protein; 357 AA.
XX AAR33440;
XX 17-DEC-2001 (revised)
DT 06-JUL-1993 (first entry)
XX Ornithine cyclodeaminase Ach5 from Ti plasmid Ach5.
XX mu-crystallins; drug targeting; nervous acting drugs; CNS; neural;
KW neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
KW neuromuscular junctions; memory agents; Alzheimers disease;
KW CNS depressants; CNS stimulators; tranquilizers; muscle relaxants;
KW antispasmodics; analgesics; anesthetics; anticonvulsants;
KW antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
KW hypnotics.
XX Agrobacterium tumefaciens.
OS USN7844304-N.
XX 01-JAN-1993.
PD 28-FEB-1992; 92US-00844304.
XX 28-FEB-1992; 92US-00844304.
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
PA Wistow G, Kim R;
XX WPI; 1993-093573/11.
DR

XX New mu-crystalline proteins - having ornithine cyclo-deaminase activity,
 PT used in diagnosis and treatment of disorders in ornithine metabolism.
 XX
 XX Disclosure; Page 34; 60pp; English.
 XX
 CC This sequence represents ornithine cyclodeaminase (OCD) from
 CC Agrobacterium Ti plasmid Ach5. It shows approximately 30% homology with
 CC the kangaroo eye lens protein mu-crystallin. (Note: Revised entry
 CC submitted to correct the patent number format of US Government-owned NTIS
 CC applications to prevent clashes with ongoing US granted patent numbers.
 CC For further information please visit the Derwent web site at
 CC www.derwent.com/dwpi/updates/ntis_us.html.)
 XX
 XX Sequence 357 AA;

Query Match 3.3%; Score 7; DB 2; Length 357;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 FGVLSDV 185
 |||||
 DB 96 FGVLSDV 102

RESULT 50
 ABU35415
 ID ABU35415 standard; protein; 368 AA.
 XX
 AC ABU35415;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #20942.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Moraxella catarrhalis.

XX WO200277183-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 XX
 XX 06-SEP-2001; 2001US-00948993.
 XX
 XX 25-OCT-2001; 2001US-0342923P.
 XX
 XX 08-FEB-2002; 2002US-00072851.
 XX
 XX 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI: 2003-029926/02.
 XX
 XX N-PSDB; ACA39285.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 25; SEQ ID NO 63339; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 XX the 6213 antisense sequences given in the specification where expression
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX (1) a vector comprising a promoter operably linked to the nucleic acid
 XX encoding a polypeptide whose expression is inhibited by the antisense
 XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX polypeptide or its fragment whose expression is inhibited by the
 XX antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs; or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 368 AA;

Query Match 3.3%; Score 7; DB 6; Length 368;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 PTHRVIS 133
 |||||
 DB 79 PTHRVIS 85

RESULT 51
 ABU40439
 ID ABU40439 standard; protein; 372 AA.
 XX
 AC ABU40439;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #25966.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Pseudomonas putida.
 XX
 XX WO200277183-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 XX
 XX 06-SEP-2001; 2001US-00948993.
 XX
 XX 25-OCT-2001; 2001US-0342923P.
 XX
 XX 08-FEB-2002; 2002US-00072851.
 XX
 XX 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI: 2003-029926/02.
 XX
 XX N-PSDB; ACA44309.

XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.

DR WPI; 2003-505284/47.
 DR N-PSDB; ADB09305.
 XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.
 XX
 PS Claim 33; SEQ ID NO 3246; 1019pp; English.
 CC The present invention describes an isolated polynucleotide (I) of
 CC Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.
 CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC expression vector comprising the novel isolated polynucleotide (I), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Alloiooccus otitidis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Alloiooccus
 CC otitidis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting Alloiooccus
 CC otitidis. The present sequence represents an Alloiooccus otitidis
 CC antigen protein from the present invention.
 XX
 SQ Sequence 425 AA;
 Query Match 3.3%; Score 7; DB 6; Length 425;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 138 FVVALSL 144
 DB 291 FVVALSL 297
 RESULT 54
 ABU23589
 ID ABU23589 standard; protein; 430 AA.
 XX
 AC ABU23589;
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #9116.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Clostridium acetobutylicum.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923F.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699F.
 XX

PA (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Chlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA27459.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 51513; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 430 AA;
 Query Match 3.3%; Score 7; DB 6; Length 430;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 22 YDKITEE 28
 DB 192 YDKITEE 198
 RESULT 55
 ABB54464
 ID ABB54464 standard; protein; 434 AA.
 XX
 AC ABB54464;
 DT 29-AUG-2003 (revised)
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein pbuX.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 OS Lactococcus lactis; IL1403.
 XX
 PN FR2807446-A1.

XX 12-OCT-2001.
 XX 11-APR-2000; 2000FR-00004630.
 XX 11-APR-2000; 2000FR-00004630.
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX Belotline A, Sorokine A, Renault P, Ehrlich SD;
 XX WPI; 2002-043418/06.
 XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species.
 XX Claim 6; SEQ ID NO 1166; 2504pp; French.
 XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (AB953300-AB955621). The nucleic
 CC acid sequence is useful in the detection and/or amplification of nucleic
 CC acid sequence, particularly to identify Lactococcus lactis or related
 CC species. The proteins of the invention are useful for the biosynthesis or
 CC biodegradation of a composition of interest. The invention helps research
 CC in lactic bacteria, particularly useful in the production of yogurt and
 CC cheese. Note: The sequence data for this patent is based on equivalent
 CC patent WO20017734 (published 18-OCT-2001) which is available in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 CC standardise OS field)
 XX SQ Sequence 434 AA;
 Query Match 3.3%; Score 7; DB 5; Length 434;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LLIAAVA 11
 Db 372 LLIAAVA 378
 |||||
 |||||
 RESULT 56
 ABU24395
 ID ABU24395 standard; protein; 453 AA.
 XX AC ABU24395;
 XX 19-JUN-2003 (first entry)
 XX Protein encoded by Prokaryotic essential gene #9922.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Clostridium botulinum.
 XX WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 XX 25-OCT-2001; 2001US-0342923P.
 XX 08-FEB-2002; 2002US-00072851.
 XX 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
 XX

DR WPI; 2003-029926/02.
 XX N-PSDB; ACA28265.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 52319; 1766pp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 453 AA;
 Query Match 3.3%; Score 7; DB 6; Length 453;
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 117 DLAYKLG 123
 Db 25 DLAYKLG 31
 |||||
 |||||
 RESULT 57
 ABB63535
 ID ABB63535 standard; protein; 460 AA.
 XX AC ABB63535;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 17397.
 XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
 XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 XX

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 52924; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 468 AA;

Query Match 3.3%; Score 7; DB 6; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 DLAYKLG 123

Db 41 DLAYKLG 47

RESULT 60

ID ADE59897 standard; protein; 468 AA.

XX ADE59897;

XX 29-JAN-2004 (first entry)

XX Human Protein P52757, SEQ ID NO 5793.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.

XX Homo sapiens.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

DR GENBANK; P52757.

XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 468 AA;

Query Match 3.3%; Score 7; DB 7; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 IEARGLK 82

Db 301 IEARGLK 307

RESULT 61

ID ABB92783 standard; protein; 474 AA.

XX ABB92783;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1994.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP009892.

XX Example 1; Fig 4; 62pp; English.

PS The invention relates to an isolated L-arabinose isomerase active enzyme

CC (1) derived from a *Thermoanaerobacter* sp. or its active fragment. The

CC enzyme is useful for converting an aldose into a ketose which involves

CC contacting the aldose with (1), and keeping the reaction under conditions

CC where at least 1% by weight of the aldose is converted. The aldose is

CC chosen from L-arabinose, D-galactose and D-fucose, and the conversion

CC reaction takes place at 60plusOC. The method further comprises the step

CC of hydrolyzing lactose to obtain D-galactose by use of a lactase -active

CC enzyme, where D-galactose is converted to D-tagatose. The lactase is

CC immobilized and the isomerase enzyme preparation is immobilized, and

CC where the lactose hydrolysis and the aldose isomerization takes place in

CC the same reactor. The lactase-active enzyme is preferably beta-

CC glycosidase. Sequences ABP71431-441 represent L-arabinose isomerase

CC (araA) enzymes from various species used for comparison studies with a T.

CC matrahnii araA enzyme

XX Sequence 498 AA;

SO

Query Match 3.3%; Score 7; DB 6; Length 498;

Best Local Similarity 100.0%; Pred. No. 3.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGGLSIL 175

Db 90 IGGLSIL 96

|||||

RESULT 64

ABB62629

ID ABB62629 standard; protein; 511 AA.

XX

AC ABB62629;

XX

XX 26-MAR-2002 (first entry)

DT

XX Drosophila melanogaster polypeptide SEQ ID NO 14679.

DE

XX Drosophila; developmental biology; cell signalling; insecticide;

XW pharmaceutical.

KW

XX Drosophila melanogaster.

OS

XX WO2001.71042-A2.

PN

XX

XX 27-SEP-2001.

PD

XX

XX 23-MAR-2001; 2001WO-US009231.

PF

XX

XX 23-MAR-2000; 2000US-0191637P.

PR

XX

XX 11-JUL-2000; 2000US-00614150.

PR

XX

PA (BEKE) PE CORP NY.

PI

XX Venter JC, Adams M, Li PWD, Myers EW;

PI

XX WPI; 2001-656860/75.

DR

XX N-PSDB; ABL06732.

DR

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

PT

XX

PS Disclosure; SEQ ID NO 14679; 2lpp + Sequence Listing; English.

PS

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SO Sequence 511 AA;

Query Match 3.3%; Score 7; DB 4; Length 511;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 KVLAPAF 206

Db 129 KVLAPAF 135

|||||

RESULT 65

ADB09310

ID ADB09310 standard; protein; 521 AA.

XX

AC ADB09310;

XX

XX 20-NOV-2003 (first entry)

DT

XX

DE Alloicoccus otitis antigenic protein SEQ ID NO:3250.

DE

XX Alloicoccus otitis; antigenic protein; immunogenic; immunisation;

XW gene therapy; Gram-positive bacterium; infection.

KW

XX Alloicoccus otitis.

OS

XX WO2003048304-A2.

PN

XX

XX 12-JUN-2003.

PD

XX

XX 25-NOV-2002; 2002WO-US036123.

PF

XX

XX 29-NOV-2001; 2001US-0333777P.

PR

XX

XX 18-NOV-2002; 2002US-0426742P.

PR

XX

PA (AMHP) WYETH HOLDINGS CORP.

PI

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

PI

XX WPI; 2003-505284/47.

DR

XX N-PSDB; ADB09309.

DR

XX

XX New Alloicoccus otitis polynucleotides and polypeptides, useful for

PT treating and diagnosing diseases, drug screening assays and monitoring of

PT effects during drug clinical trials.

PT

XX

PS Claim 33; SEQ ID NO 3250; 1019pp; English.

PS

XX The present invention describes an isolated polynucleotide (I) of

CC Alloicoccus otitis genomic DNA, which encodes an antigenic protein.

CC Alloicoccus otitis is a Gram-positive bacterium. Also described: (1)

CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an

CC expression vector comprising the novel isolated polynucleotide (I), its

CC complement, degenerate variant or fragment; (3) a genetically engineered

CC host cell, transfected, transformed or infected with the vector of (2);

CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic

CC composition comprising the polypeptide, its complement, biological

CC equivalent or fragment, or the polynucleotide that is comprised in the

CC expression vector; (6) a pharmaceutical composition comprising the

CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array

CC of the polypeptides of (1), their biological equivalent or fragment; (8)

CC immunising against Alloicoccus otitis by administering to a host the

CC immunogenic composition; (9) detecting and/or identifying Alloicoccus

CC otitis in the biological sample; (10) a kit comprising a container

CC containing the novel polynucleotide, its degenerate variant or fragment,

CC or the antibody of (4); and (11) producing a polypeptide by culturing the

CC genetically engineered host cell under conditions suitable to produce the

CC polypeptide from the culture. (I) can be used in gene therapy. The

CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting *Alloccoccus*
 CC otitis. The present sequence represents an *Alloccoccus* otitis
 CC antigen protein from the present invention.

XX Sequence 521 AA;

Query Match 3.3%; Score 7; DB 6; Length 521;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 FVVALSL 144

Db 387 FVVALSL 393

RESULT 66

ABJ39024

ID ABJ39024 standard; protein; 535 AA.

XX AC ABJ39024;

DT 17-OCT-2003 (first entry)

DE Mouse class II cytokine receptor SJ2368.

XX Class II cytokine receptor; SJ2368; autoimmune; inflammatory; cytostatic;
 KW allergic disease; septicemia; tumour; immunosuppressive; antiallergic;
 KW antinflammatory; mouse.

XX Mus sp.

OS WO2003031620-A1.

PN 17-APR-2003.

XX 02-OCT-2002; 2002WO-JP010280.

PR 02-OCT-2001; 2001JP-00306851.

PR 12-JUL-2002; 2002JP-00204385.

XX (MOCH) MOCHIDA PHARM CO LTD.

PA (KAZU-) KAZUSA DNA RES INST.

PI Ohara O, Nagase T, Katou Y, Takahashi T, Ohkawa K, Shirakawa K;

DR WPI; 2003-381719/36.

DR N-PSDB; ABT43585.

XX Class II cytokine receptor SJ2368 and regulators of its activity and
 PT expression for treatment and diagnosis of autoimmune, inflammatory and
 PT allergic diseases and tumours.

PS Claim 2; Page 159-162; 188pp; Japanese.

XX This invention relates to the class II cytokine receptor gene SJ2368 and
 CC the encoded protein, derived from either a mouse or human origin.

CC Agonists or antagonists of the cytokine receptor SJ2368 can be used for
 CC the treatment and diagnosis of autoimmune, inflammatory and allergic
 CC diseases, as well as for treating the effects of septicemia and for
 CC tumours. Accordingly, they can be described as having immunosuppressive,
 CC antinflammatory, antiallergic and/or cytostatic activity. This
 CC polypeptide sequence is the mouse class II cytokine receptor SJ2368 of
 CC the invention

XX Sequence 535 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLIAAVA 11

Db 237 LLIAAVA 243

RESULT 67

ADC17327

ID ADC17327 standard; protein; 535 AA.

XX AC ADC17327;

DT 18-DEC-2003 (first entry)

XX Mouse IFN-lambdaR1 (CRF2-12) protein.

XX mouse; IFN-lambdaR1; CRF2-12; interferon; IFN; alpha/beta independent;
 KW receptor; antiviral; IFN-lambda; IFN-lambda2; IFN-lambda3; IL-10R2;
 KW CRF2-4; cytostatic; virucidal; anti-microbial; hepatotropic;
 KW immunosuppressive; gene therapy; infection; inflammation; anti-cancer;
 KW autoimmune; hyperproliferative; cardiovascular; neurological disorder;
 KW murine.

XX Mus musculus.

PN WO2003066002-A2.

XX 14-AUG-2003.

XX 10-FEB-2003; 2003WO-US003942.

PR 08-FEB-2002; 2002US-0355196P.

PR 15-OCT-2002; 2002US-0418474P.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Kotenko S, Gallagher G;

WPI; 2003-731476/69.

DR N-PSDB; ADC17326.

XX Isolated nucleic acid molecule comprising a sequence that encodes the
 PT CRF2-12 polypeptide, useful for treating leukemias, myeloma, viral
 PT diseases, hepatitis, and autoimmune diseases.

PS Disclosure; Fig 13; 325pp; English.

XX This invention relates to a novel interferon (IFN) alpha/beta independent
 CC ligand receptor system that provides antiviral protection. Specifically,
 CC it refers to three genes located on chromosome 19, designated IFN-
 CC lambda1, IFN-lambda2 and IFN-lambda3 that are expressed upon viral
 CC infection. These IFN-lambda proteins activate a specific receptor complex
 CC for signalling and induction of various biological activities, it
 CC consists of a novel receptor known as IFN-lambdaR1 (or CRF2-12, a class
 CC II cytokine receptor family member) and a second subunit IL-10R2 (or CRF2-
 CC -4). As such, the IFN-lambda compositions of the present invention can be
 CC described as cytostatic, virucidal, anti-microbial, hepatotropic and
 CC immunosuppressive. Accordingly, via gene therapy routes for example, they
 CC can be used to treat various infections and inflammations, as well as
 CC being useful for anti-cancer therapy, autoimmune, hyperproliferative,
 CC cardiovascular and neurological disorders. This polypeptide sequence is
 CC the mouse CRF2-12 protein, homologous to the human CRF2-12 receptor of
 CC the invention.

XX Sequence 535 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLIAAVA 11

Db 237 LLIAAVA 243

```

RESULT 68
ABG25418
ID ABG25418 standard; protein; 581 AA.
XX AC ABG25418;
XX AC ABG25418;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #25409.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX KW Homo sapiens.
XX OS Homo sapiens.
XX PN W0200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS89605.
XX FT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 55777; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (II) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activities. The
XX CC polypeptide and polynucleotide sequences have application in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. NOTE: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 581 AA;

Query Match 3.3%; Score 7; DB 4; Length 581;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 DDAAIAAI 40
Db 289 DDAAIAAI 295
|||||
|||||

RESULT 69
ABG97496
ID ABG97496 standard; protein; 581 AA.
XX AC ABG97496;
XX AC ABG97496;
XX DT 16-DEC-2002 (first entry)
XX DE Human NOVX16 protein.
XX DE Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction;
XX KW Hodgkin disease; Von Hippel-Lindau syndrome; Alzheimer's disease; stroke;
XX KW tuberous sclerosis; hypercalcaemia; Parkinson's disease; depression;
XX KW Huntington's disease; cerebral palsy; epilepsy; Lesch-Nyhan syndrome;
XX KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain;
XX KW obesity; Crohn's disease; osteoporosis; inflammatory bowel disease;
XX KW infertility; inflammatory bowel disease; atherosclerosis; hypertension;
XX KW scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease;
XX KW asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis;
XX KW bacterial infection; parasitic infection; graft-versus-host disease;
XX KW cell differentiation; cell proliferation; haematopoiesis; wound healing;
XX KW angiogenesis.
XX OS Homo sapiens.
XX PN W0200272770-A2.
XX PD 19-SEP-2002.
XX PF 08-MAR-2002; 2002WO-US007283.
XX PR 08-MAR-2001; 2001US-0274281P.
XX PR 09-MAR-2001; 2001US-0274849P.
XX PR 12-MAR-2001; 2001US-0275235P.
XX PR 13-MAR-2001; 2001US-0275579P.
XX PR 13-MAR-2001; 2001US-0275601P.
XX PR 14-MAR-2001; 2001US-0276000P.
XX PR 20-MAR-2001; 2001US-0277239P.
XX PR 20-MAR-2001; 2001US-0277327P.
XX PR 20-MAR-2001; 2001US-0277338P.
XX PR 21-MAR-2001; 2001US-0277791P.
XX PR 22-MAR-2001; 2001US-0277833P.
XX PR 23-MAR-2001; 2001US-0278152P.
XX PR 26-MAR-2001; 2001US-0278894P.
XX PR 27-MAR-2001; 2001US-0279036P.
XX PR 28-MAR-2001; 2001US-0279344P.
XX PR 30-MAR-2001; 2001US-0280233P.
XX PR 02-APR-2001; 2001US-0280802P.
XX PR 02-MAY-2001; 2001US-0288148P.
XX PR 31-MAY-2001; 2001US-0294821P.
XX PR 31-OCT-2001; 2001US-0335302P.
XX PR 04-DEC-2001; 2001US-0338375P.
XX PR 07-MAR-2002; 2002US-00094466.
XX PA (CURA-) CURAGEN CORP.
XX PI Spytek KA, Vernet CA, Tchernev VT, Malvankar UM, Gerlach VL;
XX PI Li L, Zehrhusen BD, Patturajan M, Gusev VI, Kekuda R, Pena CEA;
XX PI Zhong M, Gangolli EA, Taupier RJ;
XX WPI; 2002-713508/77.
XX DR N-PSDB; ABS78740.
XX PT New NOVX polypeptides and polynucleotides, useful for preventing,
XX PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple
XX PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or
XX PT Parkinson's disease.
XX PS Claim 1; Page 137; 266pp; English.
XX CC The present invention relates to a new polypeptide (NOVX). The NOVX
XX CC polypeptide, nucleic acid and antibody are useful in the manufacture of a
XX CC medicament for treating a syndrome associated with a human disease,
XX CC preferably a NOVX-associated disorder. The NOVX nucleic acids,

```

CC polypeptides and antibodies are useful for treating, preventing or
 CC diagnosing diseases such as cancers, Hodgkin disease, Von Hippel-Lindau
 CC syndrome, Alzheimer's disease, stroke, tuberculous sclerosis, cerebral
 CC hypercalcaemia, Parkinson's disease, Huntington's disease, ataxia-
 CC palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-
 CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,
 CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,
 CC infertility, inflammatory bowel disease, atherosclerosis, hypertension,
 CC scleroderma, haemophilia, diabetes, pancreatitis, autoimmune disease,
 CC asthma, arthritis, immunodeficiencies, HIV, viral, bacterial or parasitic
 CC infections, or graft-versus-host disease. The nucleic acids and
 CC polypeptides may also be used as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOVX substances for use in therapeutic or
 CC diagnostic methods. The nucleic acids are further used as hybridisation
 CC probes, in chromosome mapping, tissue typing, preventive medicine, and
 CC pharmacogenomics. The present amino acid sequence represents a human NOVX
 CC protein of the invention
 CC
 CC Sequence 581 AA;

Query Match 3.3%; Score 7; DB 5; Length 581;

Best Local Similarity 100.0%; Pred. No. 4.4e+02; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATAAI 40

DB 289 DDATAAI 295

RESULT 70

ABU20412

ID ABU20412 standard; protein; 625 AA.

AC ABU20412;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #5939.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bacteroides fragilis.

XX WO200277183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA24282.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 48336; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 CC Sequence 625 AA;

Query Match 3.3%; Score 7; DB 6; Length 625;

Best Local Similarity 100.0%; Pred. No. 4.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 PAFKREL 210

DB 525 PAFKREL 531

RESULT 71

ADB80157

ID ADB80157 standard; protein; 657 AA.

XX AC ADB80157;

XX 04-DEC-2003 (first entry)

XX Mycobacterium tuberculosis nutrient starvation-inducible protein #98.

XX mycobacteria; nutrient starving condition; mycobacterial latency;

XX mycobacterial infection; vaccine.

OS Mycobacterium tuberculosis.

XX WO2003004520-A2.

XX 16-JAN-2003.

XX 04-JUL-2002; 2002WO-CB003052.

XX 04-JUL-2001; 2001GB-00016385.

PR 05-OCT-2001; 2001GB-00023993.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX James BW, Marsh P, Hampshire T;

XX WPI; 2003-210338/20.

DR N-PSDB; ADB80158.

PT New mycobacterial peptide, useful for the manufacture of a medicament for
 PT treating or preventing, or a diagnostic reagent for identifying,
 PT mycobacterial infection.

XX Claim 2; Page 329-332; 442pp; English.

XX The invention comprises the amino acid and coding sequences of
 CC mycobacterial (Mycobacterium tuberculosis) proteins which are up-
 CC regulated under nutrient starving conditions and maintain mycobacterial
 CC latency. The DNA and protein sequences of the invention are useful for
 CC the treatment, prevention and diagnosis of a mycobacterial infection. The
 CC present amino acid sequence represents a Mycobacterium tuberculosis
 CC protein of the invention.

XX Sequence 657 AA;

Query Match 3.3%; Score 7; DB 7; Length 657;

Best Local Similarity 100.0%; Pred. No. 5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLIIAAV 10

Db 324 FLIIAAV 330

RESULT 72

AAR38153

ID AAR38153 standard; protein; 752 AA.

AC AAR38153;

DT 24-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 13-OCT-1993 (first entry)

DE Acetobacter diguanylate phosphodiesterase PDEA2.

XW Cyclic diguanylate; diguanylate phosphodiesterase; diguanylate cyclase;
 XW cellulose production; cdg2 operon.

OS Gluconacetobacter xylinus.

Key Location/Qualifiers

FT Misc-difference 77 /note= "Val deduced from G3"

FT Misc-difference 212 /note= "Met deduced from TAG"

FT Misc-difference 215 /note= "Arg deduced from GC"

FT Misc-difference 232 /note= "Met deduced from ATC"

FT Misc-difference 233 /note= "Leu deduced from GTG"

FT Misc-difference 247 /note= "Gly deduced from GC"

FT Misc-difference 271 /note= "Gln deduced from CAGG"

FT Misc-difference 275 /note= "Val deduced from GCTG"

FT Misc-difference 388 /note= "Asp deduced from GAA"

FT Misc-difference 538 /note= "Ala deduced from CC"

XX W09311244-A1.

EN 10-JUN-1993.

XX 14-OCT-1992; 92WO-US008756.

XX 29-NOV-1991; 91US-00800218.

XX (WEYE) WEYERHAEUSER CO.

XX Tal R, Gelfand DH, Calhoun RD, Ben-Bassat A, Benziman M, Wong HC;
 PI WPI; 1993-197062/24.
 DR N-PSDB; AAQ43661.

XX Polynucleotide sequence from acetobacter CDG operon - encodes cyclic
 PT di:guanosine mono:phosphate degradation enzymes e.g. 3-phosphodiesterase
 PT isozyme.

XX Claim 5; Page 80-83; 98pp; English.

CC The amino acid sequence of protein PDEA2 was deduced from the 1st. open
 CC reading frame of the cdg2 operon. The protein is a diguanylate
 CC phosphodiesterase A, i.e. it enzymatically cleaves a single
 CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG. See also
 CC AAR38154. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24-
 CC OCT-2003 to Standardise OS field)

SQ Sequence 752 AA;

Query Match 3.3%; Score 7; DB 2; Length 752;

Best Local Similarity 100.0%; Pred. No. 5.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEQSET 45

Db 304 AIEQSET 310

RESULT 73

AB93713

ID AB93713 standard; protein; 791 AA.

AC AB93713;

XX 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2924.

XW Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP009892.

XX 28-AUG-2001; 2001WO-EP009892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidner M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.

XX Claim 5; SEQ ID NO 2924; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as

CC herbicides
XX Sequence 791 AA;
SQ
Query Match 3.3%; Score 7; DB 5; Length 791;
Best Local Similarity 100.0%; Pred. No. 5.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 AIDDAIA 38
Db 299 AIDDAIA 305
RESULT 74
ADBI1976
ID ADBI1976 standard; protein; 887 AA.
XX
AC ADBI1976;
XX
DT 20-NOV-2003 (first entry)
XX
DE Alloiococcus otitis antigenic protein SEQ ID NO:5084.
XX
KW Alloiococcus otitis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection.
XX
OS Alloiococcus otitis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
PP 25-NOV-2002; 2002WO-US036123.
XX
PR 29-NOV-2001; 2001US-0333777P.
PR 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP) WYETH HOLDINGS CORP.
XX
PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX
XX WPI; 2003-505284/47.
DR N-PSDB; ADBI1979.
XX
XX New Alloiococcus otitis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
PS Claim 33; SEQ ID NO 5084; 1019pp; English.
XX
The present invention describes an isolated polynucleotide (I) of
CC Alloiococcus otitis genomic DNA, which encodes an antigenic protein.
CC Alloiococcus otitis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
CC expression vector comprising the novel isolated polynucleotide (I), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloiococcus otitis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloiococcus
CC otitis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (I) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.

CC The polynucleotides are useful for expressing and detecting Alloiococcus
CC otitis. The present sequence represents an Alloiococcus otitis
CC antigen protein from the present invention.
XX
SQ Sequence 887 AA;
Query Match 3.3%; Score 7; DB 6; Length 887;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 DHADKPE 59
Db 627 DHADKPE 633
RESULT 75
ABU21679
ID ABU21679 standard; protein; 1001 AA.
XX
AC ABU21679;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #7206.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Burkholderia fungorum.
XX
PN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PP 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR N-PSDB; ACA25549.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 49603; 1766pp; English.
XX
The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation; (7) identifying a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1001 AA;

Query Match 3.3%; Score 7; DB 6; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 7.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AVSADPI 20
 DB 181 AVSADPI 187

RESULT 76

ADCO1332
 ID ADCO1332 standard; protein; 1027 AA.

XX AC ADCO1332;

XX 04-DEC-2003 (first entry)

DE Enterohaemorrhagic *E. coli* O157:H7-specific protein SEQ ID NO: 1377.

XX enterohaemorrhagic; anti-bacterial.

XX *Escherichia coli*; O157:H7.

XX JP2002355074-A.

XX 10-DEC-2002.

XX 24-JAN-2002; 2002JP-00015959.

XX 24-JAN-2001; 2001JP-00112010.

XX (UYTS-) UNIV TSUKUBA.

XX WPI; 2003-451640/43.

XX Enterohaemorrhagic *Escherichia coli* O157:H7-specific nucleic acid molecule
 PT and a polypeptide and its use, a polypeptide, a vector and a host cell.

XX Claim 3; SEQ ID NO 1377; 2067pp; Japanese.

XX The invention relates to a novel enterohaemorrhagic *Escherichia coli*
 CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
 CC has anti-bacterial activity. The polypeptide can be used in detection
 CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
 CC genome of Enterohaemorrhagic *E. coli* O157:H7 was determined. The present
 CC sequence represents an *E. coli* O157:H7-specific polypeptide of the
 CC invention.

XX Sequence 1027 AA;

Query Match 3.3%; Score 7; DB 7; Length 1027;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLIIAAV 10
 DB 135 FLIIAAV 141

RESULT 77

ABMI5822

ID ABMI5822 standard; protein; 1194 AA.

XX AC ABMI5822;

XX 26-SEP-2003 (first entry)

XX Mycobacterium tuberculosis mycobacterial antigen protein SEQ ID NO:11.

XX Mycobacterium tuberculosis; mycobacterial; antigen; infection; vaccine;

XX tuberculostatic; mycobacterial peptide; mycobacterial infection.

XX Mycobacterium tuberculosis.

XX WO2003033530-A2.

XX 24-APR-2003.

XX 14-OCT-2002; 2002WO-GB004647.

XX 12-OCT-2001; 2001GB-00024593.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX James B, Bacon J, March P;

XX WPI; 2003-393501/37.

XX N-PSDB; ACF39318.

XX New isolated mycobacterial peptide encoded by a gene that is induced or
 PT up-regulated under high oxygen tension, useful for diagnosing, treating
 PT or preventing a mycobacterial infection.

XX Claim 1; Page 81-86; 392pp; English.

XX The present invention describes an isolated mycobacterial peptide (I), or
 CC its fragment, variant or derivative encoded by a gene whose expression is
 CC induced or up-regulated during culture of a mycobacterium under
 CC continuous culture conditions of a dissolved oxygen tension of at least
 CC 30% air saturation measured at 37 plus degrees Celsius when compared with
 CC a dissolved oxygen tension of up to 10% air saturation measured at 37
 CC plus degrees Celsius. (I) has tuberculostatic activity and can be used in
 CC vaccines. The mycobacterial peptide (I) or its fragment, variant or
 CC derivative, inhibitor, antibody, attenuated mycobacterium, attenuated
 CC microbial carrier, DNA sequence, DNA plasmid, RNA sequence, or RNA vector
 CC from the present invention can be used for manufacturing a medicament for
 CC treating or preventing a mycobacterial infection. The peptide or its
 CC fragment, variant or derivative, the antibody, or a polynucleotide probe
 CC comprising at least 8 nucleotides, where the probe binds to at least a
 CC part of the gene, is useful for manufacturing a diagnostic reagent for
 CC identifying a mycobacterial infection. The present sequence represents a
 CC Mycobacterium tuberculosis mycobacterial antigen, which is used in the
 CC exemplification of the present invention

XX Sequence 1194 AA;

Query Match 3.3%; Score 7; DB 6; Length 1194;
 Best Local Similarity 100.0%; Pred. No. 8.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 IDDAIAA 39
 DB 45 IDDAIAA 51

RESULT 78

ADCO1365

ID ADCO1365 standard; protein; 1400 AA.

XX AC ADCO1365;

N-PSDB; ACC60993.

New isolated protein complexes useful for diagnosing a disease or disorder, or as a target for an active agent of a pharmaceutical, preferably a drug target in the treatment or prevention of disease or disorder.

Disclosure; SEQ ID NO 767; 17pp + Sequence Listing; English.

The invention relates to multiprotein complexes from eukaryotes. Proteins of the invention and DNA sequences encoding them are given in records ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are obtainable by using a protein as a bait and isolating the set of proteins which is attached thereto from cells. Such protein complexes may comprise up to 30 distinct proteins. Protein complexes of the invention are useful for diagnosing a disease or disorder, or as a target for an active agent of a pharmaceutical, preferably a drug target in the treatment or prevention of a disease or disorder. Note: The sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office. The complete document is available on CD-ROM

Sequence 2000 AA;

Query March 3.3%; Score 7; DB 6; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 177 PFGVL 183
| | | | |
Ddb 1588 PFGVL 1594

RESULT 80

ID ABB62719 standard; protein; 2836 AA.

XX ABB62719;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 14949.

DE Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

NN 27-SEP-2001.

PP 23-MAR-2001; 2001WO-US009231.

FF 23-MAR-2000; 2000US-0191637P.

RR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

FI WPI; 2001-656860/75.

XX N-PSDB; ABL06822.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

PP Disclosure; SEQ ID NO 14949; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2836 AA;
Query Match 3.3%; Score 7; DB 4; Length 2836;
Best Local Similarity 100.0%; Pred. No. 1.8e-03; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
QY 72 AMENIERA 78
DB 1376 AMENIERA 1382
|||||
RESULT 81
AAB23790
ID AAB23790 standard; peptide; 8 AA.
XX AC AAB23790;
XX DT 12-JAN-2001 (first entry)
XX DE Voltage sensitive cell adhesion molecule synthetic peptide PA22-2.
XX KW Voltage sensitive cell; VSC; adhesion molecule; binding molecule; neuron;
XX KW neurite promoting domain; hybrid electrical device; electrical junction;
XX KW transistor; artificial chemical synapse.
XX OS Synthetic.
XX PN WO200051191-A1.
XX PD 31-AUG-2000.
XX PF 22-FEB-2000; 2000MO-IL000112.
XX PR 22-FEB-1999; 99US-0121237P.
XX PS (VISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX PA Yitzchaik S, Shappir J, Spira ME;
XX WIPI; 2001-015536/02.
XX
PT A hybrid electrical device with biological components that provide an
PT artificial chemical synapse.
XX
PS Example 6; Page 13; 48pp; English.
XX
CC The present invention describes a hybrid electrical device with
CC biological components that provide an artificial chemical synapse. An
CC electrical junction (i) between one transistor and at least 1 voltage-
CC sensitive cell (VSC) characterised by at least 1 of the features selected
CC from: (i) voltage transfer between the transistor and the VSC is
CC bidirectional; (ii) there is no DC bias between the transistor and the
CC solution containing the VSC; (iii) the VSC is anchored to the external
CC surface of the transistor by binding moieties, optionally through spacers
CC ; and/or (iv) the voltage transfer between the membrane of the VSC and
CC the external surface of the transistor, and between the external surface
CC of the transistor and the membrane of the VSC is mediated by hyper-
CC polarisable chromophores. As an artificial chemical synapse i.e. a
CC junction between a cell (which secretes an agent) and a transistor
CC bearing receptors for the agent. Binding of the agent to the receptor
CC changes an electrical property of the transistor. The present sequence
CC represents a synthetic peptide, designated PA 22-2, which can be used as
CC a VSC adhesion molecule, where the VSC is a neuron, which is used in an
CC example from the present invention
XX

SQ Sequence 8 AA;
Query Match 2.8%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 VAVSAD 18
DB 2 VAVSAD 7
|||||
RESULT 82
ABG70880
ID ABG70880 standard; peptide; 8 AA.
XX AC ABG70880;
XX DT 29-AUG-2003 (revised)
XX DT 29-NOV-2002 (first entry)
XX DE R. rhodochrous nitrile hydratase beta subunit peptide #2.
XX KW Nitrile hydratase; beta subunit; enzyme; amide production;
XX KW cyanide resistance; 2-hydroxy-4-methylthiobutyronitrile;
XX KW 2-hydroxy-4-methylthiobutyroamide; alpha-hydroxy amide.
XX OS Rhodococcus rhodochrous; strain J1.
XX PN WO200270717-A2.
XX PD 12-SEP-2002.
XX PF 01-MAR-2002; 2002WO-JP001912.
XX PR 02-MAR-2001; 2001JP-00059023.
XX PR 24-JAN-2002; 2002JP-00016222.
XX PA (DAIL) DAICEL CHEM IND LTD.
XX PI Nagasawa T, Matsuyama A;
XX WIPI; 2002-707100/76.
XX
PT Novel nitrile hydratase which acts on nitrile group of nitrile compound,
PT hydrates nitrile group and converts it to amide group, useful for
PT producing amides using alpha-hydroxy nitrile as starting material.
XX
PS Example 22; Page 55; 96pp; English.
XX
CC The invention relates to a nitrile hydratase which acts on the nitrile
CC group of a nitrile compound, hydrates the nitrile group and converts it
CC to an amide group, and which is resistant to cyanide, or a protein
CC complex consisting of the alpha and beta subunits which act on a nitrile
CC group of a nitrile compound, hydrate the nitrile group and convert it to
CC an amide group, and use 2-hydroxy-4-methylthiobutyronitrile as a
CC substrate to produce 2-hydroxy-4-methylthiobutyroamide. Also included are
CC the polynucleotides encoding the alpha and beta subunits, recombinant
CC vector containing either or both polynucleotides, a transformed host
CC containing the vector, producing an amide (comprising contacting a
CC nitrile compound with the hydratase or a microorganism producing it, or
CC their processed products and recovering the amide produced) and
CC stabilising the activity of nitrile hydratase in the presence of nitrile
CC compounds, comprising contacting the nitrile hydratase with divalent
CC metal cations. The nitrile hydratase, complex or transformed host is
CC useful for producing an amide which involves contacting a nitrile
CC compound such as alpha-hydroxy- nitrile with the enzymatically active
CC material and recovering the amide (preferably alpha-hydroxy amide)
CC produced. The alpha-hydroxynitrile is represented by formula F1, and the
CC product alpha-hydroxy amide is represented by formula F3 (both detailed
CC in the specification). The present sequence represents a peptide from
CC Rhodococcus rhodochrous strain J1 nitrile hydratase beta subunit used to
CC design a PCR primer for isolation of the DNA encoding Rhodococcus sp. Cr4
CC nitrile hydratase. (Updated on 29-AUG-2003 to standardise OS field)
CC

```
XX SQ Sequence 8 AA;
Query Match 2.8%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.1e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NVVNI 169
DB 2 NVVNI 7

RESULT 83
AAR70305
ID AAR70305 standard; peptide; 10 AA.
XX AC AAR70305;
XX DT 14-NOV-1995 (first entry)
XX DE Laminin (LM) binding peptide, LM-2.
XX KW Laminin; LM; soluble; receptor; cell-adhesion.
XX OS Synthetic.
XX PN US5399667-A.
XX PD 21-MAR-1995.
XX PF 05-MAR-1993; 93US-00029333.
XX PR 05-MAR-1993; 93US-00029333.
XX PA (UNITW) UNIV WASHINGTON.
XX PI Kosfeld MD, Frazier WA;
XX WPI; 1995-130736/17.
XX Thrombospondin receptor binding peptide(s) - comprise tri:peptide
sequence Val-Val-Met (VVM).
XX Disclosure; Fig 4; 25pp; English.
XX AAR70302, AAR70303, and AAR70305 are synthetic peptides which are capable
of binding the laminin binding domain. These peptides are useful in cell
binding assays
XX SQ Sequence 10 AA;
Query Match 2.8%; Score 6; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 4 VAVSAD 9

RESULT 84
ABG67586
ID ABG67586 standard; peptide; 10 AA.
XX AC ABG67586;
XX DT 07-OCT-2002 (first entry)
XX DE Human ADPI tryptic digest peptide #295.
XX KW Human; Alzheimer's disease; AD; brain tissue; ADF; ADPI;
KW Alzheimer's disease-associated feature; neuroprotective;
KW Alzheimer's disease-associated protein isoform; nootropic;

KW ADPI tryptic digest peptide.
XX OS Homo sapiens.
XX PN WO200246767-A2.
XX PD 13-JUN-2002.
XX PF 29-NOV-2001; 2001WO-GB005289.
XX PR 08-DEC-2000; 2000US-0254431P.
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX PI Herath HMC, Parekh RB, Rohlf C;
XX WPI; 2002-508575/54.
XX Screening, diagnosis or prognosis of Alzheimer's disease in subject,
comprises detecting Alzheimer disease-associated features or Alzheimer
disease-associated protein isoforms in brain tissue from the subject.
XX Claim 7; Page 66; 427pp; English.
XX The present invention relates to methods and compositions for the
screening, diagnosis or prognosis of Alzheimer's disease (AD) in a
subject. The method comprises analysing a sample of brain tissue from a
subject by 2D electrophoresis to generate a 2D array of Alzheimer's
disease-associated features (ADFs), whose relative abundance correlates
with the presence, absence, stage or severity of AD and comparing the
abundance of each feature with the abundance of that chosen feature in
brain tissue from persons free from AD. The invention also describes in
Alzheimer's disease-associated protein isoforms (ADPis) detectable in
brain tissue. The methods and compositions of the invention are useful
for the screening, diagnosis or prognosis of AD in a subject, for
determining the stage or severity of AD in a subject, for identifying a
subject at risk of developing AD, or for monitoring the effect of therapy
administered to a subject having AD. Antibodies capable of binding to
ADPis are useful for treating or preventing AD, and for determining the
efficacy of a given treatment regime. An agent that modulates the
activity of ADPI is useful in the manufacture of a medicament for the
treatment or prevention of AD in a subject. ABG67292-ABG68038 represent
human ADPI tryptic digest peptides
XX SQ Sequence 10 AA;
Query Match 2.8%; Score 6; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 HLLIGV 107
DB 1 HLLIGV 6

RESULT 85
ABU58121
ID ABU58121 standard; peptide; 10 AA.
XX AC ABU58121;
XX DT 11-APR-2003 (first entry)
XX DE Laminin cell binding domain peptide LM22-2.
XX KW Thrombospondin-1; mouse; rat; human; chicken; tumour cell migration;
cell adhesion; platelet aggregation; wound healing; laminin; TS-1.
XX OS Synthetic.
XX PN US6469138-B1.
XX PD 22-OCT-2002.
```

XX 21-FEB-1995; 95US-00391820.
 XX 05-MAR-1993; 93US-00029333.
 XX (UNIW) UNIV WASHINGTON.
 XX Frazier WA, Kosfeld MD;
 XX WPI; 2003-196751/19.
 XX Novel peptide that binds to thrombospondin 1 receptor, useful as probe
 XX sequence in computer search of all available databases.
 XX Disclosure; Fig 4; 25pp; English.
 XX This invention relates to a novel peptide that binds to the
 XX thrombospondin 1 (TS-1) receptor; Thrombospondin 1 is a protein which
 XX acts to promote cell adhesion and is involved in platelet aggregation,
 XX wound healing and tumour cell migration. Also disclosed in the invention
 XX is laminin peptides which have been shown to be able to substitute for TS
 XX -1 peptides. Preferably the peptides of the invention share a tripeptide
 XX motif VVM and are useful as probe sequences in a computer search of all
 XX available databases for similar receptor binding motif peptides. The
 XX present sequence represents a thrombospondin-1 (TS-1) or laminin cell
 XX binding domain peptide of the invention
 XX SQ Sequence 10 AA;
 Query Match 2.8%; Score 6; DB 6; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 VAVSAD 18
 Db 4 VAVSAD 9
 RESULT 86
 ADA23705
 ID ADA23705 standard; peptide; 10 AA.
 AC ADA23705;
 XX 20-NOV-2003 (first entry)
 DT Alzheimer's disease-associated protein isoform tryptic peptide #314.
 DE human; Alzheimer's disease; vascular dementia; Lewy body dementia;
 XX schizophrenia; Parkinson's disease; multiple sclerosis; depression;
 KW Alzheimer's disease-associated protein isoform; ADPI.
 XX Homo sapiens.
 OS US2003064411-A1.
 PN 03-APR-2003.
 XX 10-DEC-2001; 2001US-00014340.
 XX 08-DEC-2000; 2000US-0254431P.
 XX (HERA/) HERATH H M A C.
 XX (PARE/) PAREKH R B.
 XX (ROHL/) ROHLFF C.
 XX Herath HMAc, Parekh RB, Rohlf C;
 XX WPI; 2003-540784/51.
 XX Screening, diagnosis or prognosis of Alzheimer's disease in subject,
 PT involves analyzing test sample of brain tissue from subject, and
 PT comparing feature in test sample with that of person(s) free from

PT Alzheimer's disease.
 XX Disclosure; SEQ ID NO 314; 115pp; English.
 XX The invention relates to a method of screening or diagnosing Alzheimer's
 XX disease in a subject. The method is useful for screening, diagnosis or
 XX prognosis of Alzheimer's disease in a subject for determining the stage
 XX of severity of Alzheimer's disease in a subject, for identifying a
 XX subject at risk of developing Alzheimer's disease, or for monitoring the
 XX effect of therapy administered to a subject having Alzheimer's disease.
 XX The method is also useful in treating vascular dementia, Lewy body
 XX dementia, schizophrenia, Parkinson's disease, multiple sclerosis or
 XX depression. The inventive method identifies sensitive and specific
 XX biomarkers for the diagnosis of Alzheimer's disease in living subjects.
 XX It provides therapeutic agents for Alzheimer's disease that works
 XX quickly, potentially, specifically with fewer side effects. The present
 XX sequence represents the amino acid sequence of a Alzheimer's disease-
 XX associated protein isoform tryptic peptide.
 XX SQ Sequence 10 AA;
 Query Match 2.8%; Score 6; DB 6; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 102 HLLIGV 107
 Db 1 HLLIGV 6
 RESULT 87
 AAW69442
 ID AAW69442 standard; protein; 11 AA.
 XX AAW69442;
 AC 10-DEC-1998 (first entry)
 DT Rat phospholipase A1 fragment.
 XX Phospholipase A1; phosphatidyl-serine hydrolysis; serine phospholipid;
 KW PLA1; enzyme; rat.
 XX Rattus sp.
 OS Key Location/Qualifiers
 FH Misc-difference 10 /note= "unspecified amino acid"
 FT JP10201479-A.
 XX 04-AUG-1998.
 XX 23-JAN-1997; 97JP-00024269.
 XX 23-JAN-1997; 97JP-00024269.
 XX (TORA) TORAY IND INC.
 XX WPI; 1998-474493/41.
 XX Phospholipase A1 and nucleic acid coding for it - having substrate
 XX specificity against serine phospholipid.
 XX Example 1; Page 5; 13pp; Japanese.
 XX This sequence is a fragment of the rat phospholipase A1 (PLA1) of the
 XX invention. PLA1 is capable of hydrolysing phosphatidyl-serine, and has
 XX substrate specificity for serine phospholipids
 XX SQ Sequence 11 AA;
 Query Match 2.8%; Score 6; DB 2; Length 11;

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 142 LSLEIS 147
| | | | |
Db 1 LSLEIS 6

RESULT 88
AAW73718
ID AAW73718 standard; peptide; 15 AA.
XX
AC AAW73718;
XX
DT 24-MAR-1999 (first entry)
XX
DE M. tuberculosis antigen clone ORF-2 peptide 2-25.
XX
KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX infection.
XX Mycobacterium tuberculosis.
XX WO9853076-A2.
PN 26-NOV-1998.
XX
PD 20-MAY-1998; 98WO-US010514.
XX
PF 20-MAY-1997; 97US-00858998.
PR 05-MAY-1998; 98US-00073009.
XX
XX (CORI-) CORIXA CORP.
XX
PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX WPI; 1999-045315/04.
XX
XX New isolated Mycobacterium tuberculosis antigens - used to develop
PT products for the prevention, treatment and diagnosis of tuberculosis
PT infection.
XX
PS Example 1; Page 64; 104pp; English.
XX
XX This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans);
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
XX Sequence 15 AA;
Qy Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 SDVLTJA 188
| | | | |
Db 8 SDVLTJA 13

RESULT 90
AAW73826

Db 3 SDVLTJA 8

RESULT 89
AAW73716
ID AAW73716 standard; peptide; 15 AA.
XX
AC AAW73716;
XX
DT 24-MAR-1999 (first entry)
XX
DE M. tuberculosis antigen clone ORF-2 peptide 2-23.
XX
KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX infection.
XX Mycobacterium tuberculosis.
XX WO9853076-A2.
PN 26-NOV-1998.
XX
PD 20-MAY-1998; 98WO-US010514.
XX
PF 20-MAY-1997; 97US-00858998.
PR 05-MAY-1998; 98US-00073009.
XX
XX (CORI-) CORIXA CORP.
XX
PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX WPI; 1999-045315/04.
XX
XX New isolated Mycobacterium tuberculosis antigens - used to develop
PT products for the prevention, treatment and diagnosis of tuberculosis
PT infection.
XX
PS Example 1; Page 63; 104pp; English.
XX
XX This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans);
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
XX Sequence 15 AA;
Qy Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 SDVLTJA 188
| | | | |
Db 8 SDVLTJA 13

RESULT 90
AAW73826

AAW73826 standard; peptide; 15 AA.
AAW73826;
24-MAR-1999 (first entry)
M. tuberculosis antigen clone ORF-2 peptide 2-23.
Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
infection.
Mycobacterium tuberculosis.
WO9853075-A2.
26-NOV-1998.
20-MAY-1998; 98WO-US010407.
20-MAY-1997; 97US-00859381.
05-MAY-1998; 98US-00073010.
(CORI-) CORIXA CORP.
Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
WPI; 1999-045314/04.
Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
useful for immunisation against M. tuberculosis infection to treat or
prevent tuberculosis, and in diagnosis of tuberculosis.
Example 1; Page 62; 100pp; English.
This sequence represents an immunogenic portion of a Mycobacterium
tuberculosis antigen of the invention. The polypeptides are useful for
immunotherapy to treat or prevent tuberculosis (especially in humans);
e.g. they can be included with an acceptable carrier in pharmaceutical
compositions or included in vaccines, and administered to induce
protective immunity in a patient against M. tuberculosis. Tuberculosis is
a chronic, infectious disease generally caused by M. tuberculosis
infection, and if left untreated typically results in serious
complications and death. Fusion proteins containing the antigen, or DNA
molecules can similarly be included with an acceptable carrier in
pharmaceutical compositions or in vaccines and administered as above. The
polypeptides are also useful for diagnosis of tuberculosis, by contacting
dermal cells with at least one polypeptide and detecting an immune
response (especially induration) on the patient's skin. Inhibiting the
spread of tuberculosis requires vaccination and accurate diagnosis, since
antibiotic therapy may not be effective due to the existence of an
asymptomatic but contagious stage and to patient non-compliance. The
polypeptides overcome concerns of safety and efficacy of current
vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
lack of sensitivity and specificity of existing diagnostic techniques
Sequence 15 AA;
Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 SDVLT A 188
Dd 8 SDVLT A 13
|||||
RESULT 91
AAW73828
ID AAW73828 standard; peptide; 15 AA.
XX
AC AAW73828;
XX
XX 24-MAR-1999 (first entry)
DT

XX DE M. tuberculosis antigen clone ORF-2 peptide 2-25.
XX AC
XX KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX KW infection.
XX OS Mycobacterium tuberculosis.
XX PN WO9853075-A2.
XX PD 26-NOV-1998.
XX PF 20-MAY-1998; 98WO-US010407.
XX PR 20-MAY-1997; 97US-00859381.
XX PR 05-MAY-1998; 98US-00073010.
XX PA (CORI-) CORIXA CORP.
XX PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX DR WPI; 1999-045314/04.
XX PS Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
XX PT useful for immunisation against M. tuberculosis infection to treat or
XX PT prevent tuberculosis, and in diagnosis of tuberculosis.
XX CC Example 1; Page 63; 100pp; English.
XX CC This sequence represents an immunogenic portion of a Mycobacterium
XX CC tuberculosis antigen of the invention. The polypeptides are useful for
XX CC immunotherapy to treat or prevent tuberculosis (especially in humans);
XX CC e.g. they can be included with an acceptable carrier in pharmaceutical
XX CC compositions or included in vaccines, and administered to induce
XX CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
XX CC a chronic, infectious disease generally caused by M. tuberculosis
XX CC infection, and if left untreated typically results in serious
XX CC complications and death. Fusion proteins containing the antigen, or DNA
XX CC molecules can similarly be included with an acceptable carrier in
XX CC pharmaceutical compositions or in vaccines and administered as above. The
XX CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
XX CC dermal cells with at least one polypeptide and detecting an immune
XX CC response (especially induration) on the patient's skin. Inhibiting the
XX CC spread of tuberculosis requires vaccination and accurate diagnosis, since
XX CC antibiotic therapy may not be effective due to the existence of an
XX CC asymptomatic but contagious stage and to patient non-compliance. The
XX CC polypeptides overcome concerns of safety and efficacy of current
XX CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
XX CC lack of sensitivity and specificity of existing diagnostic techniques
XX CC Sequence 15 AA;
XX SQ
Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 SDVLT A 188
Dd 3 SDVLT A 8
|||||
RESULT 92
AAU08183
ID AAU08183 standard; peptide; 15 AA.
XX
XX AC AAU08183;
XX
XX DT 17-DEC-2001 (first entry)
XX
XX DE Mycobacterium tuberculosis overlapping peptide ORF-2-25.
XX KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX KW immunostimulant; mutant; mutein; ORF-2-25.

```

XX OS Mycobacterium tuberculosis.
XX OS Synthetic.
XX FN WO200162893-A2.
XX PD 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US005992.
XX XX 25-FEB-2000; 2000US-0185037P.
XX PR 08-AUG-2000; 2000US-0223828P.
XX XX (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Skeiky Y, Overdale P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX PT response to and inhibiting development of a Mycobacterium infection.
XX PS Example 1; Page 134; 161pp; English.
XX CC The present invention relates to the isolation of Mycobacterium
XX CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
XX CC encoding them. The invention describes compounds and methods for the
XX CC diagnosis of tuberculosis or for inducing protective immunity against
XX CC tuberculosis. The compounds comprise at least one immunogenic portion of
XX CC one or more Mycobacterium proteins and nucleic acid molecules encoding
XX CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
XX CC encoding them can be used in diagnostic kits for the detection of
XX CC Mycobacterium infection in patients and biological samples. The compounds
XX CC of the invention and antibodies directed against the Mycobacterium
XX CC proteins may be used in vaccines for immunisation against Mycobacterium
XX CC infections. The nucleic acids encoding the Mycobacterium proteins may be
XX CC used in gene therapy. AAU08159-AAU08188 represent peptides (ORF-2-1 to
XX CC ORF-2-30) which overlap to the open reading frame ORF-2
XX SQ Sequence 15 AA;
XX Query Match 2.8%; Score 6; DB 4; Length 15;
XX Best Local Similarity 100.0%; Pred.No. 1.7e+02;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 183 SDVLTA 188
XX DB |||||
XX 3 SDVLTA 8
XX RESULT 93
XX AAU08181
XX ID AAU08181 standard; peptide; 15 AA.
XX AC AAU08181;
XX DT 17-DEC-2001 (first entry)
XX DE Mycobacterium tuberculosis overlapping peptide ORF-2-23.
XX XX Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX KW immunostimulant; mutant; mutein; ORF-2-23.
XX OS Mycobacterium tuberculosis.
XX OS Synthetic.
XX PN WO200162893-A2.
XX XX 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US005992.
XX PR 08-AUG-2000; 2000US-0185037P.
XX XX (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Skeiky Y, Overdale P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX PT response to and inhibiting development of a Mycobacterium infection.
XX PS Example 1; Page 134; 161pp; English.
XX CC The present invention relates to the isolation of Mycobacterium
XX CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
XX CC encoding them. The invention describes compounds and methods for the
XX CC diagnosis of tuberculosis or for inducing protective immunity against
XX CC tuberculosis. The compounds comprise at least one immunogenic portion of
XX CC one or more Mycobacterium proteins and nucleic acid molecules encoding
XX CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
XX CC encoding them can be used in diagnostic kits for the detection of
XX CC Mycobacterium infection in patients and biological samples. The compounds
XX CC of the invention and antibodies directed against the Mycobacterium
XX CC proteins may be used in vaccines for immunisation against Mycobacterium
XX CC infections. The nucleic acids encoding the Mycobacterium proteins may be
XX CC used in gene therapy. AAU08159-AAU08188 represent peptides (ORF-2-1 to
XX CC ORF-2-30) which overlap to the open reading frame ORF-2
XX SQ Sequence 15 AA;
XX Query Match 2.8%; Score 6; DB 4; Length 15;
XX Best Local Similarity 100.0%; Pred.No. 1.7e+02;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 183 SDVLTA 188
XX DB |||||
XX 3 SDVLTA 8
XX RESULT 94
XX AAU08181
XX ID AAR25844 standard; peptide; 19 AA.
XX AC AAR25844;
XX DT 25-MAR-2003 (revised)
XX DT 20-JAN-1993 (first entry)
XX DE Laminin peptide.
XX XX Tissue plasminogen activator; tPA; blood; clot; embolism.
XX XX Homo sapiens.
XX PN WO9211866-A1.
XX PD 23-JUL-1992.
XX XX 14-JAN-1992; 92WO-US000353.
XX PF 14-JAN-1991; 91US-00640616.
XX PR (UYDU-) UNIV DUKE.
XX PA Pizzo SV, Stack S;
XX PI WPI; 1992-268389/32.
XX XX Use of laminin fragment and tissue plasminogen activator - for treating
XX PT and preventing blood clots and embolisms.
XX XX

```


PS Claim 1; Page 15; 17pp; English.

XX The peptide is a laminin fragment comprising residues 2091-2108 of human

CC laminin with an additional Cys residue at the N-terminus. The peptide is

CC capable of activating tissue plasminogen activator and may used in a

CC compen. with tPA for the prevention and treatment of blood clots and

CC embolisms. See also AAR25845-50. (Updated on 25-MAR-2003 to correct PN

CC field.)

XX Sequence 19 AA;

XX Query Match 2.8%; Score 6; DB 2; Length 19;

XX Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSAD 18

Db 13 VAVSAD 18

RESULT 95

AAR44034

ID AAR44034 standard; peptide; 19 AA.

XX AC AAR44034;

XX 25-MAR-2003 (revised)

DT 03-DEC-1993 (first entry)

XX Laminin-derived peptide for Tc labelling, contains IKVAV.

XX integrin; cytoadhesiveness; mast cell; in vivo tumour localisation;

KW lung imaging; cell attachment; Technetium label; emphysema; cancer;

KW pulmonary disease; metal ion binding domain.

XX Synthetic.

XX Key Location/Qualifiers

FT Domain 11..15

FT /label= biological-function_domain

FT /note= "for lung imaging"

XX WO9312819-A1.

XX 08-JUL-1993.

XX 31-DEC-1992; 92WO-US011334.

XX 03-JAN-1992; 92US-00816476.

PR 03-JAN-1992; 92US-00816477.

PR 20-FEB-1992; 92US-00840077.

PR 30-DEC-1992; 92US-00998820.

PR 30-DEC-1992; 92US-00998910.

XX (RHOM-) RHOMED INC.

XX Rhodes BA, Zamora PO;

XX WPI; 1993-227063/28.

XX Metal labelled peptide(s) contg. binding and medically useful domains -

PT used in diagnosis and therapy of e.g. thrombus, cancer, infection,

PT inflammation, are also opt. combined to antibody.

XX Example 5; Page 33; 61pp; English.

XX This laminin-derived peptide contains the pentapeptide motif IKVAV, an

CC adhesive sequence from the laminin A-chain which is known to increase in

CC vitro adhesiveness of mast cells. The sequence is a preferred biological

CC function domain for peptides of the invention. In addition to a

CC biological function domain, the peptides contain a metal ion binding

CC domain (the N-terminal Cys residue in AAR44034) and are labelled with a

CC metal ion such as Technetium (esp. 99mTc). Peptides comprising IKVAV are

CC useful for lung imaging, e.g. to localise primary or metastatic cancerous

CC tumours (photon-abundant) or to monitor emphysema or fibrosis (photon-

CC deficient). (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 19 AA;

XX Query Match 2.8%; Score 6; DB 2; Length 19;

XX Best Local Similarity 100.0%; Pred. No. 2e+02;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSAD 18

Db 13 VAVSAD 18

RESULT 96

AAR92741

ID AAR92741 standard; peptide; 19 AA.

XX AC AAR92741;

XX 03-SEP-1996 (first entry)

DE IKVAV-containing sequence, for controlling cell distribution.

XX Control; distribution; bioartificial organ; BAO; cellular attachment;

KW neurotransmitter; hormone; cytokine; growth factor; enzyme.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 11..15

FT /note= "Claimed core peptide"

XX WO9602646-A2.

XX 01-FEB-1996.

XX 20-JUL-1995; 95WO-US009281.

PR 20-JUL-1994; 94US-00279773.

PR 09-MAY-1995; 95US-00432698.

XX (CYTO-) CYTOTHERAPEUTICS INC.

XX Schinstine M, Shoichet MS, Gentile FT, Hamman JP, Holland LM;

PI Cain BM, Doherty EJ, Winn SR, Aebischer P;

XX WPI; 1996-105908/11.

DR Controlling distribution of cells in bio-artificial organs - e.g. by

PT treatment of cells, or growth surfaces, to inhibit proliferation, promote

PT differentiation or modulate adhesion, for in vivo prodn. of hormones,

PT neuro-transmitter(s) etc.

XX Claim 22; Page 71; 84pp; English.

XX The sequences given in AAR92739-41 are peptides which were used in the

CC method of the invention to control the distribution of cells within a

CC bioartificial organ (BAO). These peptides have been particularly useful

CC in promoting cellular attachment. These peptides are pref. bound to the

CC membrane of the BAO which is a biocompatible, permeable jacket. These

CC peptides act to control the distribution of the core of living cells

CC included in the BAO after in vivo implantation. BAO are used

CC therapeutically to produce e.g. neurotransmitters, hormones, cytokines,

CC growth factors, enzymes, etc

XX Sequence 19 AA;

XX Query Match 2.8%; Score 6; DB 2; Length 19;

XX Best Local Similarity 100.0%; Pred. No. 2e+02;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FH Key Location/Qualifiers
 FT Region 11..15
 XX /note= "Specifically claimed in Claim 22"
 XX US2001043923-A1.
 XX 22-NOV-2001.
 XX 07-MAR-2001; 2001US-00801237.
 XX 23-MAY-1995; 95US-00447997.
 XX (SCHI/) SCHINSTINE M.
 XX (SHOI/) SHOICHET M. S.
 XX (GENT/) GENTILE F. T.
 XX (HAMW/) HAMMANG J. P.
 XX (HOLL/) HOLLAND L. M.
 XX (CAIN/) CAIN B. M.
 XX (DOHE/) DOHERTY E. J.
 XX (WINN/) WINN S. R.
 XX (AEBI/) AEBISCHER P.
 XX Schinstine M, Shoichet M, Gentile FT, Hammang JP, Holland LM;
 XX Cain BM, Doherty EJ, Winn SR, Aebischer P;
 XX WPI; 2002-082383/11.
 XX Controlling distribution of cells within a bioartificial organ useful for
 XX delivering a biologically active molecule involves exposing the cells to
 XX a treatment that inhibits or promotes cell differentiation.
 XX Disclosure; Page 10; 27pp; English.
 XX The present invention relates to a method for controlling distribution of
 XX cells within a bioartificial organ. The method involves exposing the
 XX cells to a treatment that inhibits cell proliferation or promotes cell
 XX differentiation. The treatment is effective upon in vivo implantation in
 XX a host. The treatment of the method involves suspending the cells in a
 XX hydrogel matrix within the bioartificial organ, the matrix being
 XX derivatised with a proliferation-inhibiting or differentiation-inducing
 XX peptide. The present sequence is one such peptide, which also promotes
 XX cellular attachment. The present peptide recognises a cell surface
 XX receptor molecule involved in cell adhesion
 XX Sequence 19 AA;
 XX Query Match 2.8%; Score 6; DB 5; Length 19;
 XX Best Local Similarity 100.0%; Pred. No. 2e+02;
 XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 VAVSAD 18
 DB 13 VAVSAD 18
 RESULT 100
 ABU90762
 ID ABU90762 standard; peptide; 19 AA.
 XX AC ABU90762;
 XX 11-JUL-2003 (first entry)
 XX Peptide #6 useful in cell-adhesive pattern for a biomedical device.
 XX Biomedical device; chemical surface pattern; substrate; prepattern;
 XX applied molecular self-assembly system; prefabricated pattern;
 XX cell-adhesive pattern; ligand; cell type; cell membrane receptor;
 XX analyte determination; biological sensing platform;
 XX bioanalytical sensing platform.
 XX Unidentified.
 XX

PN WO2003023401-A1.
 XX 20-MAR-2003.
 XX 12-SEP-2001; 2001WO-CH000548.
 XX 12-SEP-2001; 2001WO-CH000548.
 XX (EIDG-) EIDGENOESSISCHE TECH HOCHSCHULE ZUERICH.
 XX Textor M, Michel R, Voerces J, Hubbell JA, Lussi J;
 XX WPI; 2003-381450/36.
 XX Chemical surface patterns containing device for bioanalytical sensing
 XX platform, includes prepatterns of regions, and consecutively applied
 XX molecular self-assembly systems.
 XX Claim 73; Page 49; 69pp; English.
 XX The present invention relates to a device with chemical surface patterns
 XX with biochemical or biological relevance on substrates with prepatterns
 XX of at least two types of regions (alpha, beta), where at least two
 XX different, consecutively applied molecular self-assembly systems are used
 XX such that at least one of the applied assembly systems is specific to one
 XX type of the prefabricated patterns. The prefabricated patterns may
 XX comprise cell-adhesive patterns that contain specific ligands such as
 XX peptides, proteins and antibodies that are used to interact with specific
 XX cell type(s) or specific cell membrane receptor(s). The invention also
 XX discloses a method for the simultaneous qualitative or quantitative
 XX determination of analytes in sample(s) brought into contact with the
 XX measurement areas on a biological sensing platform. The device is useful
 XX for a bioanalytical sensing platform. ABU90757-ABU90875 represent
 XX peptides that may be used in cell-adhesive patterns for a device of the
 XX invention
 XX Sequence 19 AA;
 XX Query Match 2.8%; Score 6; DB 6; Length 19;
 XX Best Local Similarity 100.0%; Pred. No. 2e+02;
 XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 VAVSAD 18
 DB 13 VAVSAD 18
 Search completed: August 6, 2004, 16:08:23
 Job time : 58 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 16:07:23 ; Search time 18 seconds
(without alignments)
610.907 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 213
Sequence: 1 MKKFLIAAVFVAVSADPI.....VRKENTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	213	100.0	213	4	US-08-553-336A-7
2	25	11.7	215	3	US-08-462-778-2
3	25	11.7	215	4	US-08-553-336A-2
4	7	3.3	148	4	US-09-328-352-4312
5	7	3.3	204	1	US-08-292-945-8
6	7	3.3	204	1	US-08-252-073A-8
7	7	3.3	204	5	PCT-US93-1207A-8
8	7	3.3	293	4	US-09-252-991A-22215
9	7	3.3	381	4	US-08-540-236-3615
10	7	3.3	455	4	US-09-489-039A-10535
11	7	3.3	752	1	US-08-309-512-8
12	7	3.3	752	5	PCT-US92-08756A-8
13	6	2.8	10	1	US-08-029-333-31
14	6	2.8	10	4	US-08-391-820-31
15	6	2.8	12	1	US-07-616-910-54
16	6	2.8	12	1	US-08-405-200-2
17	6	2.8	12	1	US-08-405-200-15
18	6	2.8	12	5	PCT-US91-08497-54
19	6	2.8	15	4	US-09-073-009-73
20	6	2.8	15	4	US-09-073-009-75
21	6	2.8	15	4	US-09-073-010-73
22	6	2.8	15	4	US-09-073-010-75
23	6	2.8	18	1	US-07-895-252-10
24	6	2.8	18	1	US-08-072-283B-10
25	6	2.8	19	1	US-07-998-820-2
26	6	2.8	19	2	US-08-280-646-3
27	6	2.8	19	2	US-08-445-193-3
28	6	2.8	19	2	US-08-432-698-3
29	6	2.8	19	2	US-08-447-810-3
30	6	2.8	19	2	US-08-279-773-3
31	6	2.8	19	2	US-08-747-137-56
32	6	2.8	19	4	US-08-447-997-3
33	6	2.8	19	4	US-08-801-237A-3
34	6	2.8	19	5	PCT-US95-09282-3
35	6	2.8	19	5	FCT-US95-09282-3
36	6	2.8	57	4	US-08-562-737-124
37	6	2.8	67	3	US-08-905-223-335
38	6	2.8	67	4	US-09-107-532A-6800
39	6	2.8	71	4	US-05-543-681A-5866
40	6	2.8	88	4	US-08-634-238-234
41	6	2.8	91	4	US-09-543-681A-6613
42	6	2.8	94	4	US-09-073-009-25
43	6	2.8	94	4	US-09-073-009-27
44	6	2.8	94	4	US-09-073-009-29
45	6	2.8	94	4	US-09-073-010-25
46	6	2.8	94	4	US-09-073-010-27
47	6	2.8	94	4	US-09-073-010-29
48	6	2.8	101	4	US-09-252-991A-24298
49	6	2.8	102	3	US-09-383-586-34
50	6	2.8	113	4	US-09-540-236-2114
51	6	2.8	115	6	5164490-6
52	6	2.8	116	4	US-09-562-737-125
53	6	2.8	128	4	US-09-199-637A-150
54	6	2.8	130	2	US-08-944-449-2
55	6	2.8	130	4	US-09-353-362-2
56	6	2.8	133	2	US-08-891-848-15
57	6	2.8	133	3	US-08-875-811-10
58	6	2.8	134	2	US-08-891-848-14
59	6	2.8	134	3	US-08-875-811-9
60	6	2.8	134	4	US-09-446-959-5
61	6	2.8	136	4	US-09-252-991A-25498
62	6	2.8	147	3	US-09-383-586-35
63	6	2.8	152	4	US-09-446-959-7
64	6	2.8	155	4	US-09-489-039A-11904
65	6	2.8	159	4	US-09-107-532A-4926
66	6	2.8	163	4	US-09-621-976-4550
67	6	2.8	164	4	US-09-489-039A-11750
68	6	2.8	168	4	US-09-198-452A-1213
69	6	2.8	170	1	US-08-460-739-3
70	6	2.8	170	2	US-08-483-101-11
71	6	2.8	178	4	US-09-446-959-10
72	6	2.8	179	4	US-09-107-532A-3794
73	6	2.8	180	4	US-09-543-681A-7669
74	6	2.8	183	4	US-09-543-681A-5384
75	6	2.8	186	3	US-09-230-637-36
76	6	2.8	186	4	US-09-142-530C-7
77	6	2.8	186	4	US-09-198-452A-444
78	6	2.8	195	4	US-09-134-001C-4319
79	6	2.8	201	4	US-08-529-055-21
80	6	2.8	204	4	US-09-134-000C-4079
81	6	2.8	205	4	US-09-489-039A-10204
82	6	2.8	211	4	US-09-418-710-63
83	6	2.8	214	3	US-08-861-774E-44
84	6	2.8	215	4	US-09-143-127-7
85	6	2.8	222	1	US-08-028-463-4
86	6	2.8	226	1	US-08-461-836-4
87	6	2.8	226	1	US-09-489-039A-10204
88	6	2.8	230	4	US-09-489-039A-10204
89	6	2.8	231	4	US-09-495-406-35
90	6	2.8	233	2	US-08-557-309B-40
91	6	2.8	233	3	US-08-834-306-40
92	6	2.8	233	3	US-08-993-674A-40
93	6	2.8	233	4	US-09-256-976-40
94	6	2.8	235	4	US-09-374-827-8
95	6	2.8	236	4	US-09-252-991A-28804
96	6	2.8	237	4	US-09-252-991A-29211
97	6	2.8	240	4	US-09-252-991A-28714
98	6	2.8	252	4	US-09-522-714-20
99	6	2.8	252	4	US-09-489-039A-13690
100	6	2.8	256	4	US-09-543-681A-7108

ALIGNMENTS

RESULT 1

US-08-553-336A-7
; Sequence 7, Application US/08553336A
; Patent No. 6413738

GENERAL INFORMATION:

APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: IMI-032CF2

FILING DATE: 22-JUNE-1993

ATTORNEY/AGENT INFORMATION:

NAME: Jane E. Remillard

REGISTRATION NUMBER: 38,872

REFERENCE/DOCKET NUMBER: IMI-032CF2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 213 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-553-336A-7

Query Match

Best Local Similarity 100.0%; Score 213; DB 4; Length 213;

Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKFKLLIAAVAFVAVSADPHYDKITEINKAIDDAIAIEQSEITDPMKVPDPAKFER	60
DB	1	MKFKLLIAAVAFVAVSADPHYDKITEINKAIDDAIAIEQSEITDPMKVPDPAKFER	60
QY	61	HVGIVDFKGLAMRNIEARGLKQMKROGDANVKGEEGIVKXHLIGVHDDIVSMEYDLAY	120
DB	61	HVGIVDFKGLAMRNIEARGLKQMKROGDANVKGEEGIVKXHLIGVHDDIVSMEYDLAY	120
QY	121	KLGLDHPHTTHVISIQDFVVALSLEISDEGNITMTSPFVRQFANVNVNHIGLSILDPIFG	180
DB	121	KLGLDHPHTTHVISIQDFVVALSLEISDEGNITMTSPFVRQFANVNVNHIGLSILDPIFG	180
QY	181	VLSVLTAIFQDITVRKEMTKVLAPAFKRELEKN	213
DB	181	VLSVLTAIFQDITVRKEMTKVLAPAFKRELEKN	213

RESULT 2

US-08-462-778-2
; Sequence 2, Application US/08462778
; Patent No. 6077517

GENERAL INFORMATION:

APPLICANT: Thomas, Wayne R.
APPLICANT: Chua, Kaw-Yan
TITLE OF INVENTION: Allergenic Protein and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,778
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: USSN 08/031,141
FILING DATE: 12 March 1993

ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragouras

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-227-7400

TELEFAX: 617-227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 215 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-462-778-2

Query Match

Best Local Similarity 11.7%; Score 25; DB 3; Length 215;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TSFEVRQFANVNVNHIGLSILDPIF 179

DB 155 TSFEVRQFANVNVNHIGLSILDPIF 179

RESULT 3

US-08-553-336A-2

; Sequence 2, Application US/08553336A

; Patent No. 6413738

GENERAL INFORMATION:

APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996

```
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-053-336A-2

Query Match 11.7%; Score 25; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TSFEVQFANVNVHIGLSILDPF 179
DB 155 TSFEVQFANVNVHIGLSILDPF 179

RESULT 4
US-09-328-352-4312
; Sequence 4312, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: CFC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4312
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4312

Query Match 3.3%; Score 7; DB 4; Length 148;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGGLSIL 175
DB 21 IGGLSIL 27

RESULT 5
US-08-292-945-8
; Sequence 8, Application US/08292945
; Patent No. 5585478
; GENERAL INFORMATION:
; APPLICANT: Lim, Bing
; APPLICANT: Lelias, Jean-Michel
; APPLICANT: Adra, Chaker
; APPLICANT: Ko, Jong
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30309-4530

Query Match 3.3%; Score 7; DB 4; Length 148;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGGLSIL 175
DB 21 IGGLSIL 27

RESULT 6
US-08-252-073A-8
; Sequence 8, Application US/08252073A
; Patent No. 5767073
; GENERAL INFORMATION:
; APPLICANT: Bing Lim, et al.
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center,
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,073A
; FILING DATE: 1-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/292,945
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: December 10, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: YED107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
; US-08-292-945-8

Query Match 3.3%; Score 7; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 6
US-08-252-073A-8
; Sequence 8, Application US/08252073A
; Patent No. 5767073
; GENERAL INFORMATION:
; APPLICANT: Bing Lim, et al.
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center,
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,073A
; FILING DATE: 1-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
```


; REFERENCE/DOCKET NUMBER: MED107cip
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)873-8794
; TELEFAX: (404)873-8795
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
; US-08-252-073A-8

Query Match 3.3%; Score 7; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 7

PCT-US93-12074-8
; Sequence 8, Application PC/TUS9312074
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12074
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
; PCT-US93-12074-8

Query Match 3.3%; Score 7; DB 5; Length 204;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 8

US-09-252-991A-22215
; Sequence 22215, Application US/09252991A

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22215
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-22215

Query Match 3.3%; Score 7; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AFAVSA 17
DB 110 AFAVSA 116

RESULT 9

US-09-540-236-3615
; Sequence 3615, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3615
; LENGTH: 381
; TYPE: PRT
; ORGANISM: M.catarrhalis
; US-09-540-236-3615

Query Match 3.3%; Score 7; DB 4; Length 381;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 PTHVIS 133
DB 82 PTHVIS 88

RESULT 10

US-09-489-039A-10535
; Sequence 10535, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10535
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-10535

Query Match 3.3%; Score 7; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAVAFVA 14
DB 363 AAVAFVA 369

RESULT 11

US-08-309-512-8
; Sequence 8, Application US/08309512
; Patent No. 5759828
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Ben-Ziman, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoun, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,512
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: 8145-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum

US-08-309-512-8

Query Match 3.3%; Score 7; DB 1; Length 752;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEQSET 45
DB 304 AIEQSET 310

RESULT 12

PCT-US92-08756A-8

; Sequence 8, Application PC/TUS9208756A
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Ben-Ziman, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoun, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE
; TITLE OF INVENTION: METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Limbach and Limbach
; STREET: 2001 Perry Building
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08756A
; FILING DATE: 19921014
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: WEYR 20050 USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-433-4150
; TELEFAX: 415-433-8716
; TELEX: 278356
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum

PCT-US92-08756A-8

Query Match 3.3%; Score 7; DB 5; Length 752;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEQSET 45
DB 304 AIEQSET 310

RESULT 13

US-08-029-333-31
; Sequence 31, Application US/08029333
; Patent No. 5599667
; GENERAL INFORMATION:
; APPLICANT: Frazier, William A.
; APPLICANT: Kosfeld, Minh D.
; TITLE OF INVENTION: Thrombospondin Receptor Binding Peptides
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SG
; STREET: 800 N. Lindbergh Blvd.

;; CITY: St. Louis
;; STATE: Missouri
;; COUNTRY: USA
;; ZIP: 63167
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/029,333
;; FILING DATE: 19930305
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyer, Scott J.
;; REGISTRATION NUMBER: 25,275
;; REFERENCE/DOCKET NUMBER: 07-24(982)A
;; TELEPHONE: (314)694-3117
;; TELEFAX: (314)694-5435
;; INFORMATION FOR SEQ ID NO: 31:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-029-333-31

Query Match 2.8%; Score 6; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VAVSAD 18
Db 4 VAVSAD 9

RESULT 14
US-08-391-820-31
; Sequence 31, Application US/08391820
; Patent No. 6469138
; GENERAL INFORMATION:
; APPLICANT: Frazier, William A.
; TITLE OF INVENTION: Thrombospondin Receptor Binding Peptides
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SG
; STREET: 800 N. Lindbergh Blvd.
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63167
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,820
; FILING DATE: 21-Feb-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/029,333
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyer, Scott J.
; REGISTRATION NUMBER: 25,275
; REFERENCE/DOCKET NUMBER: 07-24(982)A
; TELEPHONE: (314)694-3117
; TELEFAX: (314)694-5435
; INFORMATION FOR SEQ ID NO: 31:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-08-391-820-31

Query Match 2.8%; Score 6; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VAVSAD 18
Db 4 VAVSAD 9

RESULT 15
US-07-616-910-54
; Sequence 54, Application US/07616910
; Patent No. 5223484
; GENERAL INFORMATION:
; APPLICANT: Fleming, Patrick J.
; TITLE OF INVENTION: Peptide Which Regulates Weight Gain in
; MAMMALS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/616,910
; FILING DATE: 19901121
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5223484man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-069-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-616-910-54

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 YDLAYK 121
Db 2 YDLAYK 7

RESULT 16
US-08-405-200-2
; Sequence 2, Application US/08405200
; Patent No. 5696229
; GENERAL INFORMATION:

APPLICANT: LAURIE, Gordon W
APPLICANT: MATTER, Michelle L
APPLICANT: CHEN, Lanlin
TITLE OF INVENTION: POLYPEPTIDE WITH LAMININ CELL ADHESION
TITLE OF INVENTION: AND MORPHOGENESIS ACTIVITY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,200
FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 494-150-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-200-2

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 6 VAVSAD 11

RESULT 17
US-08-405-200-15
Sequence 15, Application US/08405200
Patent No. 5696229
GENERAL INFORMATION:
APPLICANT: LAURIE, Gordon W
APPLICANT: MATTER, Michelle L
APPLICANT: CHEN, Lanlin
TITLE OF INVENTION: POLYPEPTIDE WITH LAMININ CELL ADHESION
TITLE OF INVENTION: AND MORPHOGENESIS ACTIVITY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,200

FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 494-150-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-200-15

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 6 VAVSAD 11

RESULT 18
PCT-US91-08497-54
Sequence 54, Application PC/TUS9108497
GENERAL INFORMATION:
APPLICANT: Fleming, Patrick J.
APPLICANT: Kent, Ute M.
TITLE OF INVENTION: Peptide Which Regulates Weight Gain in Mammals
TITLE OF INVENTION: Mammals
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08497
FILING DATE: 19911121
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/616,910
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, Norman F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-069-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US91-08497-54

Query Match 2.8%; Score 6; DB 5; Length 12;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YDLAYK 121
DB 2 YDLAYK 7

RESULT 19

US-09-073-009-73
; Sequence 73, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073.009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-882-6031
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; US-09-073-009-75

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 8 SDVLTA 13

RESULT 20

US-09-073-009-75
; Sequence 75, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073.009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-882-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; US-09-073-009-75

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 3 SDVLTA 8

RESULT 21

US-09-073-010-73
; Sequence 73, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073.010
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392